



BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email [info.bmjopen@bmj.com](mailto:info.bmjopen@bmj.com)

# BMJ Open

## Cohort profile: The East London Health and Care Partnership Data Repository - Using novel integrated data to support commissioning and research

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2020-037183
Article Type:	Cohort profile
Date Submitted by the Author:	22-Jan-2020
Complete List of Authors:	<p>Ronaldson, Amy; King's College London, Institute of Psychiatry, Psychology and Neuroscience; Queen Mary University of London, Centre for Psychiatry</p> <p>Chandakas, Evangelos; NHS Tower Hamlets Clinical Commissioning Group</p> <p>Kang, Qiongwen; NHS Tower Hamlets Clinical Commissioning Group</p> <p>Brennan, Katie; NHS Tower Hamlets Clinical Commissioning Group</p> <p>Akande, Aminat; NHS Tower Hamlets Clinical Commissioning Group</p> <p>Ebyarimpa, Irene; NHS Tower Hamlets Clinical Commissioning Group</p> <p>Wyllie, Eleanor; Healthy London Partnership</p> <p>Howard, George; Healthy London Partnership</p> <p>Fradgley, Richard; East London NHS Foundation Trust</p> <p>Freestone, Mark; Barts and The London School of Medicine &amp; Dentistry</p> <p>Queen Mary University of London, Centre for Psychiatry at the Wolfson Institute of Preventive Medicine</p> <p>Bhui, Kamaldeep; Barts and The London School of Medicine &amp; Dentistry</p> <p>Queen Mary University of London, Centre for Psychiatry at the Wolfson Institute of Preventive Medicine</p>
Keywords:	<p>Health economics &lt; HEALTH SERVICES ADMINISTRATION &amp; MANAGEMENT, Health policy &lt; HEALTH SERVICES ADMINISTRATION &amp; MANAGEMENT, Organisation of health services &lt; HEALTH SERVICES ADMINISTRATION &amp; MANAGEMENT, MENTAL HEALTH</p>

SCHOLARONE™  
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

## Cohort profile: The East London Health and Care Partnership Data Repository - Using novel integrated data to support commissioning and research

Amy Ronaldson, PhD<sup>1,6\*</sup>, Evangelos Handakas, PhD<sup>2</sup>, Qiongwen Kang, MSc<sup>2</sup>, Katie Brennan, MSc<sup>2</sup>, Aminat Akande<sup>2</sup>, Irene Ebyarimpa<sup>2</sup>, Eleanor Wyllie<sup>3</sup>, George Howard<sup>3</sup>, Richard Fradgley<sup>4</sup>, Mark Freestone, PhD<sup>1,5</sup>, Kamaldeep Bhui, MD<sup>1</sup>

<sup>1</sup>Centre for Psychiatry, Wolfson Institute of Preventive Medicine, Barts & The London School of Medicine, Old Anatomy Building, Charterhouse Square, London EC1M 6BQ

<sup>2</sup>Tower Hamlets Clinical Commissioning Group, Mile End Hospital, Bancroft Road, Bethnal Green, London E1 4DG

<sup>3</sup>Healthy London Partnership, 75-77 Worship Street, Hackney, London EC2A 2DU

<sup>4</sup>East London NHS Foundation Trust, Trust Headquarters, Robert Dolan House, 9 Alie Street, London E1 8DE

<sup>5</sup>The Alan Turing Institute, British Library, London NW1 2DB

<sup>6</sup>Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, De Crespigny Park, London SE5 8AF

**Keywords:** Mental health; health services; research cohort; commissioning; health policy

**Word count:** 3,505

**Number of tables and figures:** 10

**\*Corresponding author and person to whom reprint requests should be addressed:**

Dr Amy Ronaldson, Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, De Crespigny Park, London SE5 8AF

Email: [amy.ronaldson@kcl.ac.uk](mailto:amy.ronaldson@kcl.ac.uk)

**Disclosure statement:** Dr Mark Freestone received research grants from East London NHS Foundation Trust and The Alan Turing Institute during the conduct of this study

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Funding:** Funding for this project was granted by East London NHS Foundation Trust (ELFT), the Tower Hamlets Together Board, the Health London Partnership Mental Health Transformation Board, and the City & Hackney Transformation Board. The funder had no role in the preparation of this manuscript. This report is independent research supported by the National Institute for Health Research ARC North Thames. The views expressed in this publication are those of the author(s) and not necessarily those of the National Institute for Health Research or the Department of Health and Social Care.

For peer review only

## **Abstract**

**Purpose:** The East London Health and Care Partnership (ELHCP) Data Repository was established to support commissioning decisions in London. This dataset comprises routine clinical data for the GP-registered populations of two London boroughs, Tower Hamlets and City & Hackney, and provides a rich source of demographic, clinical, and health service use data of relevance to clinicians, commissioners, researchers and policy makers. This paper describes the dataset, its representativeness, and data completeness.

**Participants:** There were 351,749 and 344,511 members of the GP-registered population in the two boroughs respectively for the financial year 2017/18. Demographic information, and prevalence data was available for nine mental health and 15 physical health conditions. Prevalence was compared with local and national data for validation. In order to illustrate the health service use data available in the dataset, emergency department use across mental health conditions was described. Information about data completeness was provided.

**Findings to date:** The ELHCP Data Repository provides a rich source of information about a relatively young, urban, ethnically diverse, population within areas of socioeconomic deprivation. Prevalence data were in line with local and national statistics with some exceptions. Physical health conditions were more common in those with mental health conditions, reflecting that co-morbidities are the norm rather than the exception. This has implications for integrated care. Mental health conditions increased emergency department visits 54% and 76% more in each borough, suggesting mental health conditions may have significant cost implications for the physical health system. The precise drivers of increased demand need further evaluation. Data completeness for risk factors (e.g. blood pressure, cholesterol) was high in patients with long-term conditions.

**Future plans:** The data are being further cleaned and evaluated using imputation, Bayesian, and economic methods. Data continue to be collected and will enable more long-term prospective analysis as data become available.

**Strengths and limitations of this study**

- This dataset will allow for the analysis of a relatively young, urban, ethnically diverse, considerably deprived population – groups that are hard to research with traditional survey methods.
- The data are linked across care settings, allowing for the evaluation of integrated interventions.
- Large amounts of data around health service utilisation and associated spend means this dataset can be used to support healthcare commissioning and health service delivery.
- Reliance on primary care data for diagnostic information means that prevalence rates can be under-estimated and certain information can be missed.
- As is the case with all routine clinical data, the dataset requires extensive cleaning to remove errors.

**Introduction**

The use of observational data is fundamental to healthcare service provision and health research, and increasingly large amounts of routinely collected clinical data are becoming available. The main advantage of these datasets are their comprehensive nature, and large patient numbers [1]. The development of these large routine datasets also allow for the application of big data analytic techniques within health research [2]. These techniques can evaluate population characteristics, identify risk factors, and allow for the development of predictive models relating to clinical outcomes [1,3]. Moreover, these datasets identify areas of high health service utilisation and high-cost patients which might support decision-making around health service delivery for commissioners, policy makers, and healthcare providers. To date, several large datasets comprising routine clinical data have been developed in the United Kingdom (UK); these are being used to inform clinical practice and healthcare delivery [4–6].

Recently, a large-scale routine clinical dataset has been developed in the Tower Hamlets borough of London, UK, comprising data from the general practitioner (GP)-registered populations of Tower Hamlets and its neighbouring borough, City & Hackney. The East London Health and Care Partnership (ELHCP) Data Repository was developed primarily to support commissioning decisions and health service delivery in East London and comprises rich information relevant to population health and clinical research. These areas of London are unique inner-city areas with diverse populations and high levels of deprivation. Tower Hamlets is home to the largest Bangladeshi community in England [7]. It also has the highest rate of

poverty, child poverty, and unemployment of any London borough [8]. What might be of interest to researchers of severe mental illness is that the borough of City & Hackney has the highest incidence of psychosis in England [9].

The aim of this cohort profile is to describe the dataset in Tower Hamlets and City & Hackney, comparing population demographics, prevalence rates of mental and physical health conditions, health service utilisation, and data completeness. We believe these data to be useful for clinicians, researchers, commissioners, and policy makers, particularly those with an interest in understanding health in urban, diverse, and deprived areas.

### *The development of the ELHCP Data Repository*

The ELHCP Data Repository was developed by Tower Hamlets Clinical Commissioning Group (CCG) (<https://www.towerhamletsccg.nhs.uk/>) and North East London Commissioning Support Unit (NEL CSU) (<http://www.nelcsu.nhs.uk/>), with funding provided by Tower Hamlets CCG and Tower Hamlets Together (THT) Vanguard (<https://www.towerhamletstogether.com/>). Development of the dataset started in 2015 and efforts to expand the dataset (e.g. to all areas in North East London) and improve data quality are ongoing.

The dataset comprises up-to-date patient level data for GP-registered populations in the London boroughs of Tower Hamlets and City & Hackney and is updated every three months. All patient data is depersonalised and linked by replacing patient NHS numbers with unique ID numbers. Data pertaining to geographic area (i.e. Lower Super Output Area (LSOA)) and other population-level data are drawn from data aggregated by the Office of National Statistics (ONS). Presently, secure data streams are in place for: acute (inpatient, outpatient, and emergency) services; community services; mental health services; substance misuse services; and primary care services (e.g. diagnostic information, prescribing). Efforts are underway to establish secure data streams for social care data, public health data (e.g. sexual health), and urgent care data (e.g. walk-in-centres).

As illustrated in Figure 1, data for acute, community, and mental health services flow to the CCG via NHS digital, The Data Services for Commissioners Regional Office (DSCRO), and local CSUs. Pseudonymisation of the data is performed by the DSCRO hosted by the CSU. The data then flows to the local CSU which formats and organises the data for use at the CCG. Primary care data flows to NEL CSU where a pseudonymisation process is carried out which



mirrors that performed by the DSCRO. These data then flow to the local CSU alongside the acute, community, and mental health data where they are added to the ELHCP Data Repository.

**Methods**

This cohort profile is a descriptive piece detailing the following:

- (i) Demographic information (e.g. age, sex, ethnicity, and deprivation)
- (ii) Prevalence rates of mental and physical health conditions
- (iii) Data relating to health service utilisation in patients with mental health conditions
- (iv) Information about data completeness

All data presented in this profile were from financial year 2017/2018.

*Population demographics*

Demographic information for these populations included age, sex, ethnicity, and deprivation index. There was also information relating to body mass index (BMI), and smoking status. Information about age, sex, and ethnicity came from primary care records. Patients were classified into nine ethnic groups: white, or not stated; Indian; Pakistani; Bangladeshi; other Asian; Black Caribbean; Black African; Chinese; other ethnic group. Deprivation was measured using the Index of Deprivation. This measure is provided for LSOAs – small areas that are focused on social homogeneity and a defined population size (1000 – 1500 residents) [10]. The ELHCP dataset allocates each LSOA in Tower Hamlets and City & Hackney to a decile of deprivation with ‘1’ being the most deprived decile. Information relating to BMI and smoking status came from primary care records.

*Mental and physical health conditions*

Primary care data were used to identify and define nine mental health conditions and 15 physical health conditions. These conditions are listed in Table 1. Variables pertaining to each condition were formulated by NEL CSU using EMIS Read Codes (v.2 and v.3). Lists of the codes used to formulate each variable are available on request. All physical health variables were deemed long-term conditions and therefore a patient who had ever had a primary care diagnosis recorded were included in the physical health cohort. This was the same for severe mental illness (SMI, i.e. schizophrenia, psychosis, bipolar disorder), personality disorder, anorexia, bulimia, dementia, and learning disabilities. As depression, low mood, and anxiety can potentially be more episodic conditions, further work was done to determine the timelines

which should be used to define these cohorts. Based on local clinical and academic advice, and available evidence [11] we decided to create two patient cohorts for each conditions – patients who had ever received a diagnosis and patients who had a diagnosis recorded in the last 12 months.

Looking at the impact of the method of definition on activity in secondary care services indicated that the preferable way to define the depression, low mood, and anxiety cohorts was to include those who had ever had a diagnosis recorded in primary care. This was partly because only looking at those who had received a diagnosis in the past 12 months meant that we omitted a considerable proportion of patients who likely had common mental health disorders (CMDs, i.e. depression and/or anxiety) from our analyses. If required, it is possible to apply defined periods of time to conditions where full prolonged remission or cure is a possibility, e.g. cancer.

Prevalence rates for each condition in both Tower Hamlets and City & Hackney were calculated from the data. In order to assess the face validity of the diagnoses, we assessed population prevalence in the dataset and compared with both the local and national population from other data sources. The prevalence rates of all 15 physical health conditions across six of the mental health conditions were also calculated in order to illustrate the bidirectional links between mental and physical health.

### *Health service utilisation*

Health service utilisation data outlined in this profile covered use of emergency services for exemplary purposes. Data relating to specific mental health conditions were described. Health service utilisation data for elective and non-elective inpatient admissions, and outpatient appointments are also available in the ELHCP Data Repository. These data come from the SUS dataset and provide number of attendances/admissions/appointments per patient per year. There is also data pertaining to associated spend (e.g. spend per activity, spend per patient per year) which is currently being validated and is therefore not presented in this paper.

### *Data completeness*

As the ELHCP Data Repository comprises routinely collected data from several different data flows there is considerable variation in completeness across variables. Completeness of certain variables appears to be better in patients with diagnoses of mental/physical health conditions compared to those with no diagnoses. We used type 2 diabetes and COPD as exemplar physical conditions, and depression and SMI as exemplar mental conditions to demonstrate how data

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

quality differs across patient groups. We provided completeness counts on several demographic and clinical variables including age, sex, deprivation level, ethnicity, smoking status, alcohol use, BMI, weight (kg), systolic blood pressure (SBP), diastolic blood pressure (DBP), and cholesterol. All variables were from primary care datasets. Alcohol use was measured using a composite variable of combined scores from the Alcohol Use Disorders Identification Test (AUDIT) and AUDIT-consumption (AUDIT-C) completed in primary care.

*Patient and public involvement*

This cohort was developed without patient involvement.

**Cohort description and findings to date**

*Population demographics*

The GP-registered populations for Tower Hamlets and City & Hackney in financial year 2017/2018 were 351,749 and 344,511 respectively. See Table 2 for a summary of population demographics for both Tower Hamlets and City & Hackney. Data for age, sex, and smoking status were recorded for the complete GP-registered populations in both East London boroughs. Deprivation index was recorded for the majority of the population (99.8% complete in both boroughs). Ethnicity data was recorded to a good level of completeness at 86.6% and 84.6% in Tower Hamlets and City & Hackney respectively. BMI data was more problematic and was recorded in the primary care records for less than a quarter of GP-registered patients (Tower Hamlets: 23.4%, City & Hackney: 24.0%).

Demographic data suggested that both Tower Hamlets and City & Hackney are ‘young’ boroughs relative to English national data[12] (median age 39 years) with mean ages of 32 (median = 31 years) and 34 years (median = 33 years) respectively. In terms of ethnicity, the majority of the GP-registered population in Tower Hamlets (38.2%) and City & Hackney (46.8%) were of white (or not stated) ethnicity. A considerable proportion of the population of Tower Hamlets were of Bangladeshi ethnicity (27.4%) and a significant amount of the City & Hackney population belonged to ‘other ethnic group’ (16.8%). Most of the GP-registered population in both East London boroughs were living in areas of high deprivation: in Tower Hamlets and City & Hackney 58.8% and 54.2% of the population respectively were in deciles 1 and 2, indicative of high deprivation. Approximately 20% of the population were recorded in primary care as being smokers. Although BMI is poorly recorded in primary care records,

mean BMI scores indicated that the GP-registered population on average were overweight (Tower Hamlets:  $26.3 \pm 8.1$ ; City & Hackney:  $27.1 \pm 8.6$ ).

### *Prevalence rates of mental and physical health conditions*

Prevalence rates for GP-coded diagnoses of mental health conditions in both Tower Hamlets and City & Hackney are provided in Figure 2. Prevalence rates were similar across both boroughs for all conditions. Comparative local and national prevalence data for each condition is provided in Table 3. The GP-coded prevalence of common mental health disorders (CMDs) was 16.8% and 17.5% in Tower Hamlets and City & Hackney respectively. This was in line with data from the latest Adult Psychiatric Morbidity Survey (APMS) which reported that 17% of the English population have a CMD[13]. However, the GP-coded prevalence of depression in Tower Hamlets (8.7%) and City & Hackney (11.4%) was higher than Public Health England (PHE) prevalence data[14] from the same areas from the same financial year (7.3%, 9.6%). This could be due to the GP-coded prevalence in this profile reflecting diagnoses of depression that were ever recorded, which likely led to somewhat inflated rates.

GP-coded prevalence of SMI (1.3%) from the ELHCP Data Repository was in line with data from the PHE Psychosis Data Report from 2016[9] which reported prevalence rates of SMI for both Tower Hamlets and City & Hackney at 1.32% and 1.36% respectively. Prevalence rates for learning disability were also in line with local and national data provided by Public Health England[14]. However, there were discrepancies between the GP-coded prevalence of personality disorder, eating disorders, and dementia in the ELHCP Data Repository and local and national PHE[14] and APMS[13] data. What this suggests is that these conditions might be under-recorded in primary care in these boroughs.

Prevalence rates for GP-coded diagnoses of physical health conditions in the ELHCP Data Repository are provided in Figure 3. Prevalence rates were similar across boroughs for most conditions. However, primary care records showed that levels of type 2 diabetes were higher in Tower Hamlets and levels of hypertension were higher in City & Hackney. Comparative local and national prevalence data for each condition is provided in Table 4. Local and national prevalence data was sourced from PHE public health profiles[14]. Overall, GP-coded prevalence of physical health conditions from the ELHCP Data Repository seems to be in line with PHE local data for the same areas. There is one notable discrepancy, however. Levels of asthma were much higher in the ELHCP Data Repository. This might be due to overdiagnosis of asthma in primary care which has been reported in both adults[15] and children[16]. Within

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

the ELHCP Data Repository ‘liver disease’ comprises many liver complications, e.g. hepatitis, fatty liver disease, alcoholic liver disease. Therefore, it was not possible to directly compare prevalence data relating to liver disease with local and national statistics as PHE provide individual prevalence rates for each liver disease subtype.

*Prevalence of physical health conditions among mental health conditions*

Mental and physical health are inextricably linked[17]. Having a mental health condition predisposes an individual to develop problems with their physical health. At the same time, living with a physical health condition can greatly impact mental health. Figure 4a and Figure 4b illustrate the prevalence of physical health conditions among mental health conditions in the ELHCP Data Repository. Mental conditions included in this analysis were depression, low mood, anxiety, SMI, and personality disorder.

Results indicate that levels of all 15 physical health conditions were increased in those with mental health conditions. The distribution of the prevalence of physical health conditions amongst mental health conditions was similar across boroughs. The most pronounced increases in physical health conditions were seen for those flagged by their GP as having depression, SMI, and personality disorder. Some striking findings included the increase in levels of type 2 diabetes and hypertension among patients flagged by their GP as having SMI. Type 2 diabetes is known to be elevated in patients with SMI[18,19]. This increase might be brought about through use of antipsychotic medications known to be both diabetogenic and linked with weight gain[20], as well as through social deprivation, poor nutrition, chronic stress, and lifestyle changes[21]. Associations between SMI and hypertension reported in the literature have been more mixed[21]. Another striking finding that occurred in both boroughs was the high prevalence of asthma for people with a primary care diagnosis of personality disorder. This may be linked to smoking which is known to be increased in patients with personality disorder[22]. Personality disorder has been linked to hypochondriasis and perceived breathlessness[23], which might also be contributing factors to higher levels of asthma diagnosis in primary care for these patients.

*A&E attendances and associated spend across mental health conditions*

We used A&E as an example of a health service utilisation outcome measure to illustrate some of the activity data available in the ELHCP Data Repository. Data relating to use of A&E across both boroughs are detailed in Table 5. Based on crude attendance rates (per 1000 population),

patients with mental health conditions attended A&E 54% and 76% more in Tower Hamlets and City & Hackney respectively compared to those without GP-recorded mental health conditions. Increased attendance rates were particularly pronounced in patients with dementia, personality disorder, and SMI.

### *Data completeness*

The ELHCP Data Repository relies on data inputted by health care provider organisations (GP practices, clinical coders in acute care settings, etc.). It is comprised of routinely collected data that can vary in completeness across variables. Examination of four example physical and mental health conditions illustrated how data completeness varies across patient groups. Table 6 details completeness on key demographic and clinical variables for patients with type 2 diabetes, COPD, depression, and SMI in Tower Hamlets. These data were all recorded in primary care databases. Data pertaining to age, sex, deprivation level, ethnicity, and smoking status were complete for all patients included in the analysis. Data relating to alcohol use and clinical factors differed between those with and without health conditions in that the data was considerably more complete for patients registered as having a condition. This improvement in completeness is likely down to the implementation of financial incentive schemes introduced as part of the UK's Quality Outcomes Framework (QOF) to improve clinical management of patients with chronic conditions[4,24]. What this illustrates is that the ELHCP Data Repository is well-suited to perform research on individual physical and mental health conditions using primary care data, where data completeness is to a high standard.

### **Strengths and limitations**

#### *Strengths*

One of the main strengths of the dataset is that it provides the opportunity to analyse a relatively young, urban, ethnically diverse, and considerably deprived population: groups that are hard to reach with traditional survey methods. Given patient records are linked across care settings, the data can be used to evaluate the impacts of integrated interventions. In this profile we have illustrated that the ELHCP Data Repository is representative in terms of the prevalence of most mental and physical health conditions compared to other sources of local and national prevalence data supporting its suitability for real-world research and decision-making. This dataset also allows for the analysis of comorbidities and multimorbidity. Moreover, there is a large amount of data available around health service utilisation and associated spend meaning



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

that the dataset can be used to understand what drives use and spend and support healthcare commissioning and health service delivery.

The dataset is in regular use by commissioners to better understand the health and care needs of the population, highlight where there is scope for improved patient care, and evaluate the impact of health and care interventions of patients and the system. This has provided commissioners, and others working on health and care system integration, with more sophisticated evidence and analysis to inform decisions about where to target improvement efforts to maximise benefit to both patients and the system. The findings presented in this cohort profile have highlighted where links between mental and physical health are most prevalent in the population and helped to shape commissioning priorities accordingly. For example, analysis highlighting mental and physical comorbidity supported conversations about expanding IAPT services within long term condition pathways, noting that it may be particularly impactful for people with type 2 diabetes, respiratory conditions, and cancer. Furthermore, over the next year Tower Hamlets is investing in a new 24/7 core fidelity crisis service, and the transformation of community mental health services to become embedded in primary care networks. These new models will support the assessment and treatment of patients at the right place at the right time. The ELHCP Data Repository described in this cohort profile will be an important tool in evaluating the impact these service changes have on patients and the healthcare system.

In terms of completeness, this paper highlighted that data were considerably more complete for patients with long-term mental or physical illness (i.e. type 2 diabetes, COPD, depression, SMI) compared to patients without. This supports the suitability of the dataset for the analysis of patients with long-term conditions.

*Limitations*

In terms of limitations associated with the dataset, in some cases prevalence data was not in line with local and national data, meaning that it is possible some conditions are under-recorded in primary care, i.e. personality disorder, eating disorders, and dementia. Also, as prevalence data relies on GP diagnoses, it is not possible to gauge disease duration in patients as the date in which a diagnosis is first recorded might not be the date the patient received the diagnosis. For example, the diagnosis might have been recorded on the day the patient registered with their GP, reflecting that date rather than the date of diagnosis. In terms of data completeness, although this is good for patients with long-term conditions, there is considerable missingness

among 'healthy' patients in the dataset. It is possible to resolve this with imputation to some degree. As is the case with all routine clinical data, the dataset requires extensive cleaning to remove duplicate entries and spurious values out of clinical range (e.g. BMI, blood pressure) prior to analysis. Another limitation involves the amount of data available. Currently, there are two financial years of data that are at an acceptable standard for research and/or decision support. New data continues to be collected and flow into this data set, so the time series available will increase over time.

### **Conclusions and future collaborations**

This dataset provides a unique opportunity to analyse a relatively young, urban, diverse, and deprived population in East London. The dataset is growing year on year and data quality should improve over time making this an invaluable resource for researchers, clinicians, commissioners, and policy makers. The data are being further cleaned and evaluated using imputation, Bayesian, and economic methods.

### **Author statement**

AR drafted the manuscript with input and critical revisions from all authors. QK, KB, EW, GH, RF, MF, and KB provided supervision throughout. KB is the principal investigator and grant holder. Analysis of the data was performed by AA, EH, and IE. All authors reviewed the paper before submission.

### **References**

- 1 Cook JA, Collins GS. The rise of big clinical databases. *Br J Surg* 2015;**102**:e93–101. doi:10.1002/bjs.9723
- 2 Bates DW, Saria S, Ohno-Machado L, *et al.* Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Aff Proj Hope* 2014;**33**:1123–31. doi:10.1377/hlthaff.2014.0041
- 3 Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst* 2014;**2**:3. doi:10.1186/2047-2501-2-3
- 4 Correa A, Hinton W, McGovern A, *et al.* Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) sentinel network: a cohort profile. *BMJ Open* 2016;**6**:e011092. doi:10.1136/bmjopen-2016-011092
- 5 Herrett E, Gallagher AM, Bhaskaran K, *et al.* Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015;**44**:827–36. doi:10.1093/ije/dyv098
- 6 Perera G, Broadbent M, Callard F, *et al.* Cohort profile of the South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLaM BRC) Case



- Register: current status and recent enhancement of an Electronic Mental Health Record-derived data resource. *BMJ Open* 2016;**6**:e008721. doi:10.1136/bmjopen-2015-008721
- 7 Tower Hamlets Council. Ethnicity in Tower Hamlets: Analysis of 2011 census data. 2013. [https://www.towerhamlets.gov.uk/Documents/Borough\\_statistics/Ward\\_profiles/Census-2011/RB-Census2011-Ethnicity-2013-01.pdf](https://www.towerhamlets.gov.uk/Documents/Borough_statistics/Ward_profiles/Census-2011/RB-Census2011-Ethnicity-2013-01.pdf) (accessed 7 Mar 2019).
  - 8 Trust for London. Poverty and inequality data for Tower Hamlets. Trust Lond. 2017./data/boroughs/tower-hamlets-poverty-and-inequality-indicators/ (accessed 3 Jul 2019).
  - 9 Public Health England. Psychosis data report. 2016;:77.
  - 10 Noble M, McLennan D, Wilkinson K, *et al*. English Indices of Deprivation. London: : Communities and Local Government 2007.
  - 11 John A, McGregor J, Fone D, *et al*. Case-finding for common mental disorders of anxiety and depression in primary care: an external validation of routinely collected data. *BMC Med Inform Decis Mak* 2016;**16**:35. doi:10.1186/s12911-016-0274-7
  - 12 Ethnicity facts and figures: Age groups. HM Government Services 2018. <https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/demographics/age-groups/latest#download-the-data> (accessed 6 Jun 2019).
  - 13 Adult Psychiatric Morbidity Survey: Survey of Mental Health and Wellbeing, England, 2014. NHS Digit. <https://digital.nhs.uk/data-and-information/publications/statistical/adult-psychiatric-morbidity-survey/adult-psychiatric-morbidity-survey-survey-of-mental-health-and-wellbeing-england-2014> (accessed 11 Jan 2019).
  - 14 Public Health England. Public Health Profiles. <https://fingertips.phe.org.uk/profile-group/mental-health> (accessed 18 Jun 2019).
  - 15 Aaron SD, Vandemheen KL, Boulet L-P, *et al*. Overdiagnosis of asthma in obese and nonobese adults. *CMAJ* 2008;**179**:1121–31. doi:10.1503/cmaj.081332
  - 16 Akker IL den, Luijn K van, Verheij T. Overdiagnosis of asthma in children in primary care: a retrospective analysis. *Br J Gen Pract* 2016;**66**:e152–7. doi:10.3399/bjgp16X683965
  - 17 Doherty AM, Gaughran F. The interface of physical and mental health. *Soc Psychiatry Psychiatr Epidemiol* 2014;**49**:673–82. doi:10.1007/s00127-014-0847-7
  - 18 Das-Munshi J, Ashworth M, Gaughran F, *et al*. Ethnicity and cardiovascular health inequalities in people with severe mental illnesses: protocol for the E-CHASM study. *Soc Psychiatry Psychiatr Epidemiol* 2016;**51**:627–38. doi:10.1007/s00127-016-1185-8
  - 19 Ward M, Druss B. The epidemiology of diabetes in psychotic disorders. *Lancet Psychiatry* 2015;**2**:431–51. doi:10.1016/S2215-0366(15)00007-3

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
- 20 Gianfrancesco FD, Grogg AL, Mahmoud RA, *et al.* Differential effects of risperidone, olanzapine, clozapine, and conventional antipsychotics on type 2 diabetes: findings from a large health plan database. *J Clin Psychiatry* 2002;**63**:920–30.
- 21 Osborn DPJ, Wright CA, Levy G, *et al.* Relative risk of diabetes, dyslipidaemia, hypertension and the metabolic syndrome in people with severe mental illnesses: systematic review and metaanalysis. *BMC Psychiatry* 2008;**8**:84. doi:10.1186/1471-244X-8-84
- 22 Zvolensky MJ, Jenkins EF, Johnson KA, *et al.* Personality Disorders and Cigarette Smoking among Adults in the United States. *J Psychiatr Res* 2011;**45**:835–41. doi:10.1016/j.jpsychires.2010.11.009
- 23 Chetta A, Gerra G, Foresi A, *et al.* Personality Profiles and Breathlessness Perception in Outpatients with Different Gradings of Asthma. *Am J Respir Crit Care Med* 1998;**157**:116–22. doi:10.1164/ajrccm.157.1.9702093
- 24 Kontopantelis E, Reeves D, Valderas JM, *et al.* Recorded quality of primary care for patients with diabetes in England before and after the introduction of a financial incentive scheme: a longitudinal observational study. *BMJ Qual Saf* 2013;**22**:53–64. doi:10.1136/bmjqs-2012-001033

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Table 1.** Mental and physical health conditions recorded in primary care records

Mental health conditions
Depression; low mood*; anxiety; severe mental illness (SMI); personality disorder; anorexia; bulimia; dementia; learning disability
Physical health conditions
Type 1 diabetes, type 2 diabetes, asthma, chronic obstructive pulmonary disorder (COPD), cancer, stroke, coronary heart disease (CHD), heart failure, atrial fibrillation, hypertension, peripheral artery disease (PAD), chronic kidney disease (CKD), liver/pancreas complications, osteoporosis, epilepsy
*low mood is often recorded in primary care in place of depression and is therefore important to consider in any analyses relating to mental health

**Table 2.** Population demographics for the GP-registered populations of Tower Hamlets and City & Hackney

	<b>Tower Hamlets</b>	<b>City &amp; Hackney</b>
<b>GP registered population (n,%)</b>	351749 (100%)	344511 (100%)
<b>Age (years) (M±SD)</b>	32 ± 17.97	34 ± 19.47
<b>Sex (n,%)</b>		
<i>Male</i>	178088 (50.63%)	168371 (48.87%)
<i>Female</i>	173656 (49.37%)	176134 (51.13%)
<b>Ethnicity (n,%)</b>		
<i>White or not stated</i>	134411 (38.21%)	161069 (46.75%)
<i>Indian</i>	9295 (2.64%)	9067 (2.63%)
<i>Pakistani</i>	3547 (1.01%)	2707 (0.79%)
<i>Bangladeshi</i>	96418 (27.41%)	5877 (1.71%)
<i>Other Asian</i>	8151 (2.32%)	6189 (1.80%)
<i>Black Caribbean</i>	4281 (1.22%)	16052 (4.66%)
<i>Black African</i>	14034 (3.99%)	30037 (8.72%)
<i>Chinese</i>	10435 (2.97%)	4137 (1.20%)
<i>Other ethnic group</i>	23934 (6.80%)	56140 (16.29%)
<b>Deprivation index (n,%)</b>		
<i>High deprivation*</i>	206723 (58.77%)	186784 (54.22%)
<i>Low deprivation**</i>	4955 (1.41%)	3134 (0.91%)
<b>Smokers (n,%)</b>	69627 (19.79%)	72371 (21.01%)
<b>BMI (kg/m<sup>2</sup>) (M±SD)</b>	26.34 ± 8.28	27.06 ± 8.58

\*High deprivation refers to the GP-registered population in decile 1 and 2

\*\*Low deprivation refers to the GP-registered population in decile 9 and 10

BMI = body mass index; M: Mean; SD: Standard deviation

Table 3. GP-coded prevalence of mental health conditions in Tower Hamlets and City & Hackney compared to local and national prevalence data (sources: APMS 2014, PHE Public Health Profiles, PHE Psychosis Data Report 2016)

Condition	GP-coded prevalence		Local prevalence		National prevalence
	<i>Tower Hamlets</i>	<i>City &amp; Hackney</i>	<i>Tower Hamlets</i>	<i>City &amp; Hackney</i>	<i>England</i>
<b>CMDs</b>	16.8%	17.5%	15.4%[14]	17%[14]	17%[13]
<b>SMI</b>	1.3%	1.3%	1.32%[9]	1.36%[9]	0.5-1.5%[9]
<b>Personality disorder</b>	0.5%	0.6%	-	-	13.7%[13]*
<b>Eating disorders</b>	0.3%	0.4%	6.9%[14]	7.4%[14]	6.9%[14]
<b>Dementia</b>	0.4%	0.4%	5.25%[14]	4.37%[14]	4.33%[14]
<b>Learning disability</b>	0.5%	0.6%	0.3%[14]	0.4%[14]	0.5%[14]

\*Percentage of sample who scored  $\geq 4$  on the Standardised Assessment of Personality – Abbreviated Scale (SAPAS)

APMS = Adult Psychiatric Morbidity Survey; CMDs = common mental health disorders; GP = general practitioner; PHE = Public Health England; SMI = severe mental illness

**Table 4.** GP-coded prevalence of physical health conditions in Tower Hamlets and City & Hackney compared to local and national prevalence data (source: PHE Public Health Profiles)

Condition	GP-coded prevalence		Local prevalence		National prevalence
	<i>Tower Hamlets</i>	<i>City &amp; Hackney</i>	<i>Tower Hamlets</i>	<i>City &amp; Hackney</i>	<i>England</i>
<b>Type 1 diabetes*</b>	0.3%	0.4%			
<b>Type 2 diabetes*</b>	5.1%	4.4%	6.7%	5.7%	6.8%
<b>Asthma</b>	8.7%	9.2%	4.3%	4.4%	5.9%
<b>COPD</b>	1.3%	1.2%	1.2%	1.1%	1.9%
<b>Cancer</b>	1.4%	1.8%	1.1%	1.4%	2.7%
<b>Stroke</b>	0.8%	0.9%	0.9%	0.7%	1.8%
<b>CHD</b>	1.6%	1.7%	1.5%	1.5%	3.1%
<b>Heart failure</b>	0.7%	0.9%	0.4%	0.7%	0.8%
<b>Atrial fibrillation</b>	0.6%	0.8%	0.6%	0.7%	1.9%
<b>Hypertension</b>	7.4%	9.3%	7.5%	9.3%	13.9%
<b>PAD**</b>	0.3%	0.4%	0.96%	1.01%	1.16%
<b>CKD</b>	2.0%	1.8%	2.5%	2.1%	4.1%
<b>Liver disease†</b>	1.9%	1.4%	-	-	-
<b>Osteoporosis</b>	0.06%	0.09%	-	-	0.6%
<b>Epilepsy</b>	0.8%	0.8%	0.4%	0.5%	0.8%

\*PHE does not provide prevalence rates for diabetes subtypes. Therefore local and national prevalence rates is for diabetes overall.

\*\*Local and national prevalence data for PAD from PHE were based on modelled estimated prevalence

†It was not possible to provide local and national prevalence rates for liver disease as data is provided separately by PHE for disease subtype (e.g. hepatitis C, fatty liver disease, alcoholic liver disease)

CHD = coronary heart disease; CKD = chronic kidney disease; COPD = chronic obstructive pulmonary disorder; GP = general practitioner; PAD = peripheral artery disease; PHE = Public Health England

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Table 5.** The impact of mental health conditions on A&E attendances, financial year 2017/2018

Accident & Emergency						
Activity (attendances)						
	Total attendances		% Share		Crude rate (per 1000 pop.)	
	TH	C&H	TH	C&H	TH	C&H
Population	159201	161641	100%	100%	452	469
All non-mental health	118272	116816	74.3%	72.3%	412	413
All mental health	40929	44825	25.7%	27.7%	635	728
CMDs	31625	34215	19.9%	21.2%	535	568
SMI	3800	4383	2.4%	2.7%	847	945
Personality disorder	2529	2914	1.6%	1.8%	1413	1467
Eating disorder	584	549	0.4%	0.3%	554	424
Dementia	1595	1471	1.0%	0.9%	1248	1075
Learning disability	796	1293	0.5%	0.8%	670	974

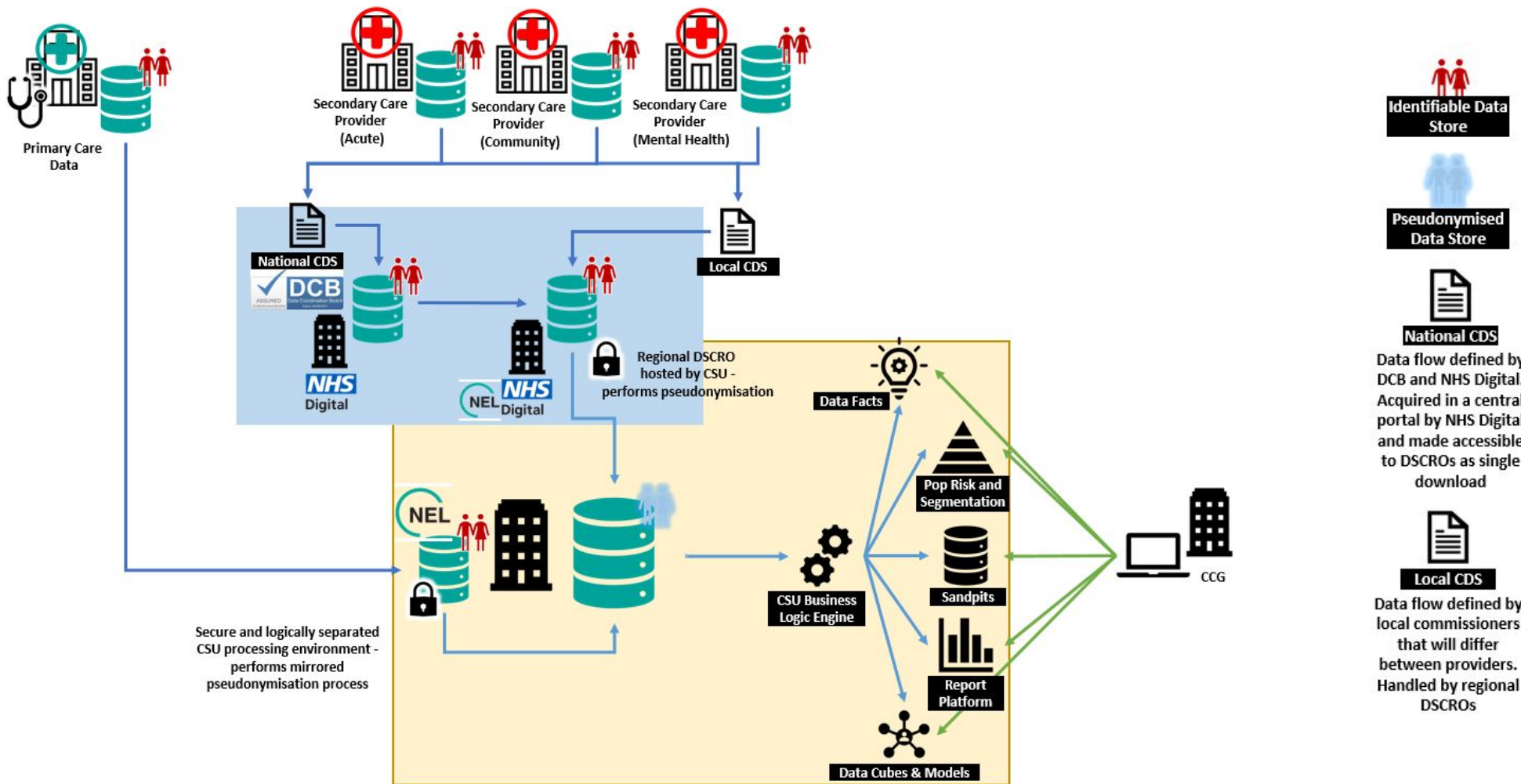
A&E = Accident & Emergency; CMDs = common mental health disorders; C&H = City & Hackney; SMI = severe mental illness; TH = Tower Hamlets

**Table 6.** ELHLP data completeness (%) across physical and mental health conditions in Tower Hamlets, financial year 2017/2018

	<b>T2DM</b>	<b>No T2DM</b>	<b>COPD</b>	<b>No COPD</b>	<b>Depression</b>	<b>No Depression</b>	<b>SMI</b>	<b>No SMI</b>
<i>Variables</i>	<i>% complete</i>	<i>% complete</i>	<i>% complete</i>	<i>% complete</i>	<i>% complete</i>	<i>% complete</i>	<i>% complete</i>	<i>% complete</i>
<b>Age</b>	100%	100%	100%	100%	100%	100%	100%	100%
<b>Sex</b>	100%	100%	100%	100%	100%	100%	100%	100%
<b>Deprivation</b>	100%	100%	100%	100%	100%	100%	100%	100%
<b>Ethnicity</b>	100%	100%	100%	100%	100%	100%	100%	100%
<b>Smoking status</b>	100%	100%	100%	100%	100%	100%	100%	100%
<b>Alcohol use</b>	94.6%	51.1%	91.8%	52.8%	73.1%	51.4%	89.2%	52.8%
<b>BMI</b>	99.3%	72.6%	99.3%	73.7%	95.4%	71.9%	97.5%	73.7%
<b>Weight</b>	99.4%	84.6%	99.5%	85.2%	96.7%	84.3%	98.2%	85.2%
<b>SBP</b>	99.6%	68.8%	99.6%	69.9%	96.4%	67.9%	98.5%	69.9%
<b>DBP</b>	99.6%	68.7%	99.6%	69.9%	96.3%	67.9%	98.5%	69.9%
<b>Cholesterol</b>	99.4%	28.8%	96.9%	31.6%	64.9%	29.3%	84.6%	37.8%

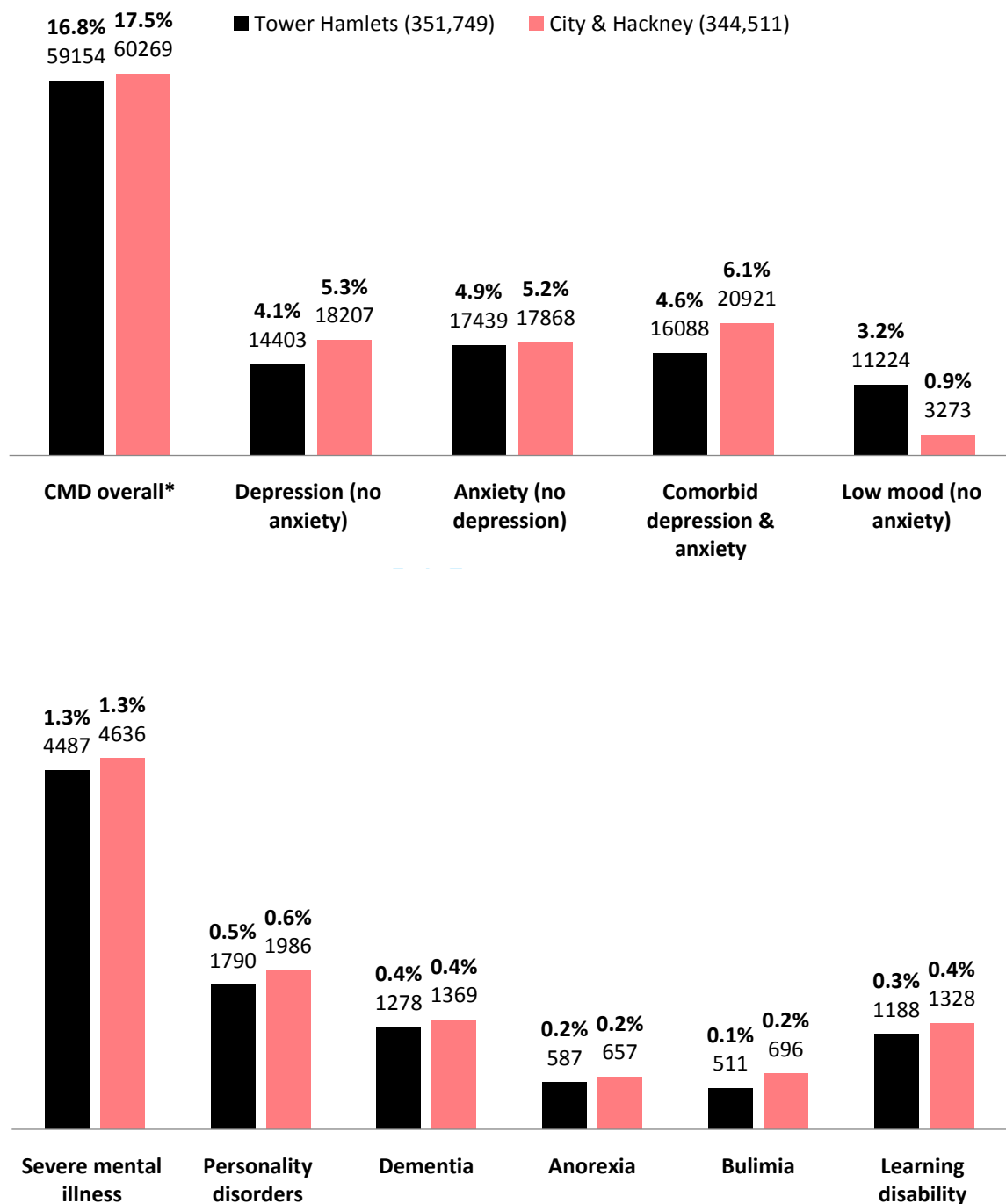
BMI = body mass index; COPD: chronic obstructive pulmonary disorder; DBP = diastolic blood pressure; ELHLP = East London Healthcare Links Project; SBP = systolic blood pressure; SMI = severe mental illness; T2DM = Type 2 diabetes mellitus



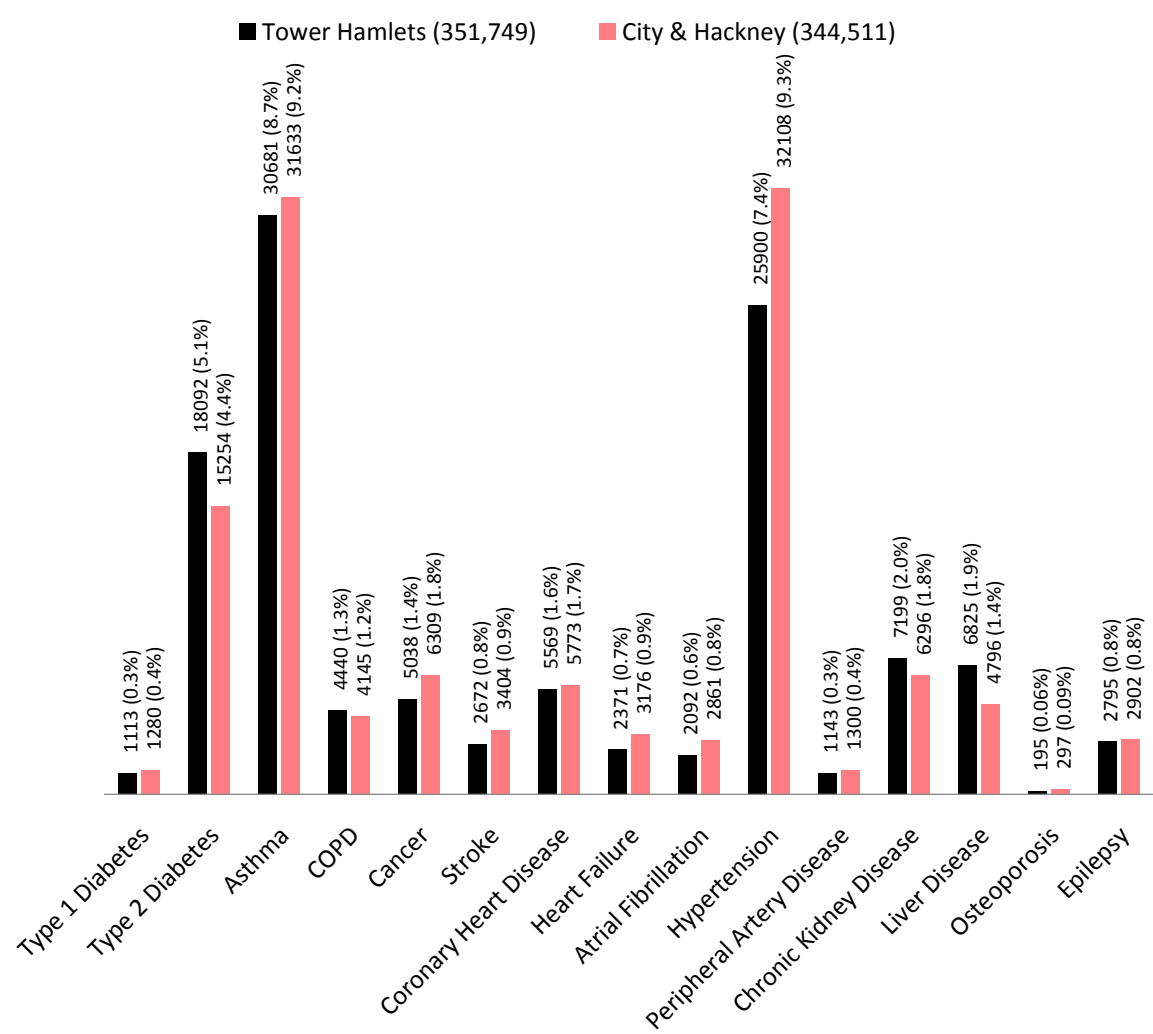


**Figure 1.** Diagram of data flows from healthcare providers to Tower Hamlets & City Hackney CCGs

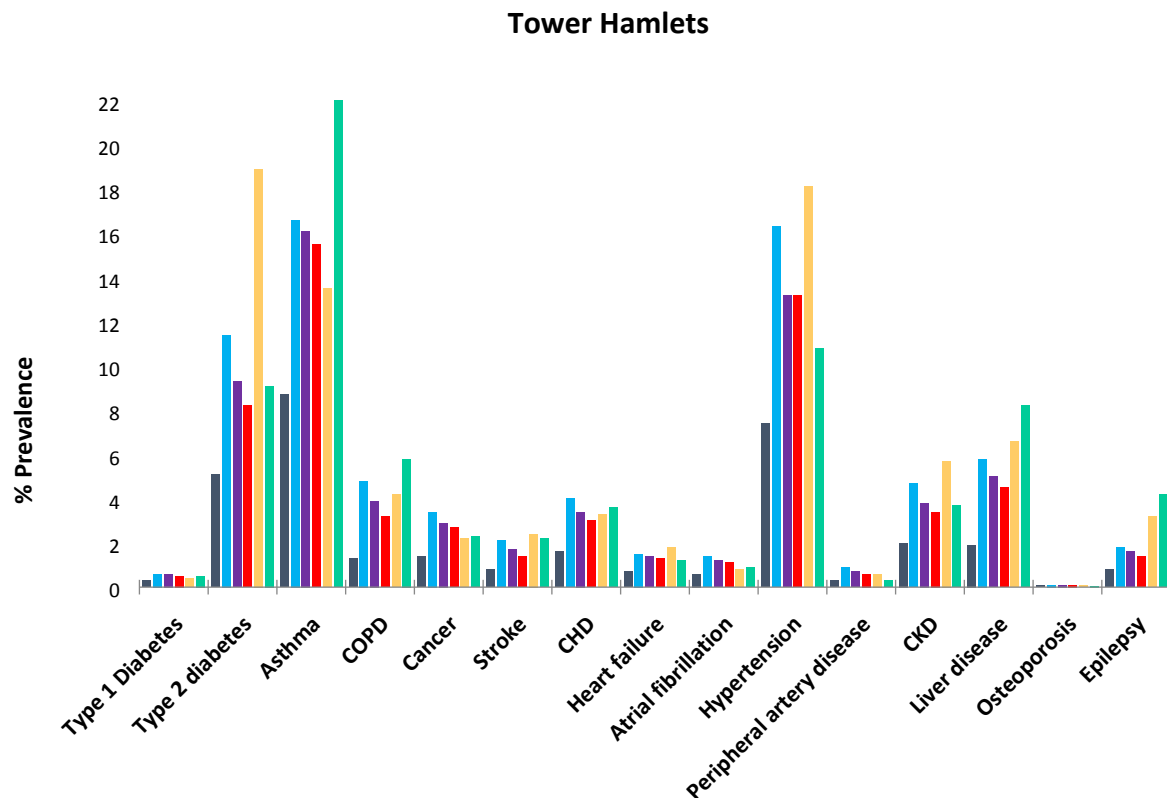
CCG = Clinical Commissioning Group, CDS = Commissioning Data Sets, CSU = Commissioning Support Unit, DCB = Data Coordination Board, DSCRO = Data Services for Commissioners Regional Office, NEL = North East London, NHS = National Health Service



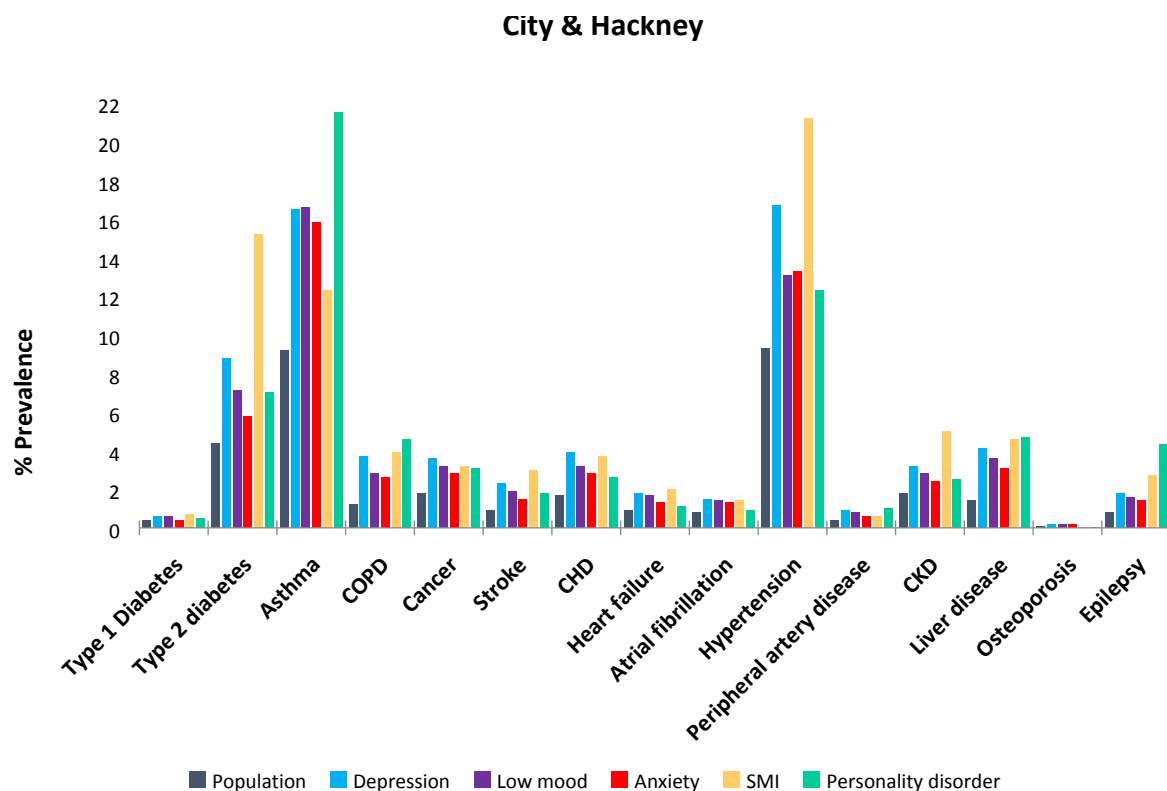
**Figure 2.** GP-coded prevalence rates of mental health conditions in Tower Hamlets and City & Hackney, financial year 2017/2018



**Figure 3.** GP-coded prevalence rates of physical health conditions in Tower Hamlets and City & Hackney, financial year 2017/2018



**Figure 4a.** Prevalence of physical health conditions among mental health conditions in Tower Hamlets



**Figure 4b.** Prevalence of physical health conditions among mental health conditions in City & Hackney

# BMJ Open

## Cohort profile: The East London Health and Care Partnership Data Repository - Using novel integrated data to support commissioning and research

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2020-037183.R1
Article Type:	Cohort profile
Date Submitted by the Author:	30-Jun-2020
Complete List of Authors:	Ronaldson, Amy; King's College London, Institute of Psychiatry, Psychology and Neuroscience; Queen Mary University of London, Centre for Psychiatry Chandakas, Evangelos; NHS Tower Hamlets Clinical Commissioning Group Kang, Qiongwen; NHS Tower Hamlets Clinical Commissioning Group Brennan, Katie; NHS Tower Hamlets Clinical Commissioning Group Akande, Aminat; NHS Tower Hamlets Clinical Commissioning Group Ebyarimpa, Irene; NHS Tower Hamlets Clinical Commissioning Group Wyllie, Eleanor; Healthy London Partnership Howard, George; Healthy London Partnership Fradgley, Richard; East London NHS Foundation Trust Freestone, Mark; Barts and The London School of Medicine & Dentistry Queen Mary University of London, Centre for Psychiatry at the Wolfson Institute of Preventive Medicine Bhui, Kamaldeep; Barts and The London School of Medicine & Dentistry Queen Mary University of London, Centre for Psychiatry at the Wolfson Institute of Preventive Medicine
<b>Primary Subject Heading</b>:	Health informatics
Secondary Subject Heading:	Health services research, Health policy, Public health, Health economics
Keywords:	Health economics < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Health policy < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, MENTAL HEALTH, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, EPIDEMIOLOGY, PRIMARY CARE

SCHOLARONE™  
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

## **Cohort profile: The East London Health and Care Partnership Data Repository - Using novel integrated data to support commissioning and research**

Amy Ronaldson, PhD<sup>1,6\*</sup>, Evangelos Chandakas, PhD<sup>2</sup>, Qiongwen Kang, MSc<sup>2</sup>, Katie Brennan, MSc<sup>2</sup>, Aminat Akande<sup>2</sup>, Irene Ebyarimpa<sup>2</sup>, Eleanor Wyllie<sup>3</sup>, George Howard<sup>3</sup>, Richard Fradgley<sup>4</sup>, Mark Freestone, PhD<sup>1,4,5</sup>, Kamaldeep Bhui, MD<sup>1</sup>

<sup>1</sup>Centre for Psychiatry, Wolfson Institute of Preventive Medicine, Barts & The London School of Medicine, Old Anatomy Building, Charterhouse Square, London EC1M 6BQ

<sup>2</sup>Tower Hamlets Clinical Commissioning Group, Mile End Hospital, Bancroft Road, Bethnal Green, London E1 4DG

<sup>3</sup>Healthy London Partnership, 75-77 Worship Street, Hackney, London EC2A 2DU

<sup>4</sup>East London NHS Foundation Trust, Trust Headquarters, Robert Dolan House, 9 Alie Street, London E1 8DE

<sup>5</sup>The Alan Turing Institute, British Library, London NW1 2DB

<sup>6</sup>Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, De Crespigny Park, London SE5 8AF

**Keywords:** Mental health; health services; research cohort; commissioning; health policy

**Word count:** 3,505

**Number of tables and figures:** 10

**\*Corresponding author and person to whom reprint requests should be addressed:**

Dr Amy Ronaldson, Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, De Crespigny Park, London SE5 8AF

Email: [amy.ronaldson@kcl.ac.uk](mailto:amy.ronaldson@kcl.ac.uk)

**Competing interest statement:** Dr Mark Freestone received research grants from East London NHS Foundation Trust and The Alan Turing Institute during the conduct of this study.

**Funding:** Funding for this project was granted by East London NHS Foundation Trust (ELFT), the Tower Hamlets Together Board, the Health London Partnership Mental Health Transformation Board, and the City & Hackney Transformation Board. The funder had no role in the preparation of this

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

manuscript. This report is independent research supported by the National Institute for Health Research  
ARC North Thames. The views expressed in this publication are those of the author(s) and not  
necessarily those of the National Institute for Health Research or the Department of Health and Social  
Care.

**Data availability statement:** The data is housed at the Tower Hamlets Clinical Commissioning Group.  
Permission to access the data will need to be requested from Tower Hamlets Clinical Commissioning  
Group.

For peer review only



## **Abstract**

**Purpose:** The East London Health and Care Partnership (ELHCP) Data Repository was established to support commissioning decisions in London. This dataset comprises routine clinical data for the GP-registered populations of two London boroughs, Tower Hamlets and City & Hackney, and provides a rich source of demographic, clinical, and health service use data of relevance to clinicians, commissioners, researchers and policy makers. This paper describes the dataset in its current form, its representativeness, and data completeness.

**Participants:** There were 351,749 and 344,511 members of the GP-registered population in the two boroughs respectively for the financial year 2017/18. Demographic information and prevalence data was available for nine mental health and 15 physical health conditions. Prevalence was compared with local and national data for validation. In order to illustrate the health service use data available in the dataset, emergency department use across mental health conditions was described. Information about data completeness was provided.

**Findings to date:** The ELHCP Data Repository provides a rich source of information about a relatively young, urban, ethnically diverse, population within areas of socioeconomic deprivation. Prevalence data were in line with local and national statistics with some exceptions. Physical health conditions were more common in those with mental health conditions, reflecting that co-morbidities are the norm rather than the exception. This has implications for integrated care. Data completeness for risk factors (e.g. blood pressure, cholesterol) was high in patients with long-term conditions.

**Future plans:** The data are being further cleaned and evaluated using imputation, Bayesian, and economic methods, principally focusing on specific cohorts, including type II diabetes, depression and personality disorder. Data continue to be collected for the foreseeable future to support commissioning decisions which will also enable more long-term prospective analysis as data become available at the end of each financial year.

**Strengths and limitations of this study**

- This dataset will allow for the analysis of a relatively young, urban, ethnically diverse, considerably deprived population – groups that are hard to research with traditional survey methods.
- The data are linked across care settings, allowing for the evaluation of integrated interventions.
- Large amounts of data around health service utilisation and associated spend means this dataset can be used to support healthcare commissioning and health service delivery.
- Reliance on primary care data for diagnostic information means that prevalence rates can be under-estimated and certain information can be missed.
- As is the case with all routine clinical data, the dataset requires extensive cleaning to remove errors.

**Introduction**

The use of observational data is fundamental to healthcare service provision and health research, and increasingly large amounts of routinely collected clinical data are becoming available. The main advantage of these datasets are their comprehensive nature, and large patient numbers [1]. The development of these large routine datasets also allow for the application of big data analytic techniques within health research [2]. These techniques can evaluate population characteristics, identify risk factors, and allow for the development of predictive models relating to clinical outcomes [1,3]. Moreover, these datasets identify areas of high health service utilisation and high-cost patients which might support decision-making around health service delivery for commissioners, policy makers, and healthcare providers. To date, several large datasets comprising routine clinical data have been developed in the United Kingdom (UK); these are being used to inform clinical practice and healthcare delivery [4–6].

Recently, a large-scale routine clinical dataset has been developed in the Tower Hamlets borough of London, UK, comprising data from the general practitioner (GP)-registered populations of Tower Hamlets and its neighbouring borough, City & Hackney. The East London Health and Care Partnership (ELHCP) Data Repository was developed primarily to support commissioning decisions and health service delivery in East London and comprises rich information relevant to population health and clinical research. These areas of London are unique inner-city areas with diverse populations and high levels of deprivation. Tower Hamlets is home to the largest Bangladeshi community in England [7]. It also has the highest rate of

poverty, child poverty, and unemployment of any London borough [8]. What might be of interest to researchers of severe mental illness is that the borough of City & Hackney has the highest incidence of psychosis in England [9].

The aim of this cohort profile is to introduce and describe the dataset in Tower Hamlets and City & Hackney, comparing population demographics, prevalence rates of mental and physical health conditions, health service utilisation, and data completeness. We believe this cohort profile might facilitate the development of other similar datasets. Moreover, these data will prove useful for clinicians, researchers, commissioners, and policy makers, particularly those with an interest in understanding health in urban, diverse, and deprived areas.

### *The development of the ELHCP Data Repository*

The ELHCP Data Repository was developed by Tower Hamlets Clinical Commissioning Group (CCG) (<https://www.towerhamletsccg.nhs.uk/>) and North East London Commissioning Support Unit (NEL CSU) (<http://www.nelcsu.nhs.uk/>), with funding provided by Tower Hamlets CCG and Tower Hamlets Together (THT) Vanguard (<https://www.towerhamletstogether.com/>). Development of the dataset started in 2015 and efforts to expand the dataset (e.g. to all areas in North East London) and improve data quality are ongoing.

The dataset comprises up-to-date patient level data for GP-registered populations in the London boroughs of Tower Hamlets and City & Hackney and is updated every three months. All patient data is depersonalised and linked by replacing patient NHS numbers with unique ID numbers. Data pertaining to geographic area (i.e. Lower Super Output Area (LSOA)) and other population-level data are drawn from data aggregated by the Office of National Statistics (ONS). Presently, secure data streams are in place for: acute (inpatient, outpatient, and emergency) services; community services; mental health services; substance misuse services; and primary care services (e.g. diagnostic information, prescribing). Efforts are underway to establish secure data streams for social care data, public health data (e.g. sexual health), and urgent care data (e.g. walk-in-centres).

As illustrated in Figure 1, data for acute, community, and mental health services flow to the CCG via NHS digital, The Data Services for Commissioners Regional Office (DSCRO), and local CSUs. Pseudonymisation of the data is performed by the DSCRO hosted by the CSU. The data then flows to the local CSU which formats and organises the data for use at the CCG.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Primary care data flows to NEL CSU where a pseudonymisation process is carried out which mirrors that performed by the DSCRO. These data then flow to the local CSU alongside the acute, community, and mental health data where they are added to the ELHCP Data Repository.

**Methods**

This cohort profile is a descriptive piece detailing the following:

- (i) Demographic information (e.g. age, sex, ethnicity, and deprivation)
- (ii) Prevalence rates of mental and physical health conditions
- (iii) Data relating to health service utilisation in patients with mental health conditions
- (iv) Information about data completeness

All data presented in this profile were from financial year 2017/2018.

*Population demographics*

Demographic information for these populations included age, sex, ethnicity, and deprivation index. There was also information relating to body mass index (BMI), and smoking status. Information about age, sex, and ethnicity came from primary care records. Patients were classified into nine ethnic groups: white, or not stated; Indian; Pakistani; Bangladeshi; other Asian; Black Caribbean; Black African; Chinese; other ethnic group. Deprivation was measured using the Index of Deprivation. This measure is provided for LSOAs – small areas that are focused on social homogeneity and a defined population size (1000 – 1500 residents) [10]. The ELHCP dataset allocates each LSOA in Tower Hamlets and City & Hackney to a decile of deprivation with ‘1’ being the most deprived decile. Information relating to BMI and smoking status came from primary care records.

*Mental and physical health conditions*

Primary care data were used to identify and define nine mental health conditions and 15 physical health conditions. These conditions are listed in Table 1. Variables pertaining to each condition were formulated by NEL CSU using EMIS Read Codes (v.2 and v.3). Lists of the codes used to formulate each variable are available on request. All physical health variables were deemed long-term conditions and therefore a patient who had ever had a primary care diagnosis recorded were included in the physical health cohort. This was the same for severe mental illness (SMI, i.e. schizophrenia, psychosis, bipolar disorder), personality disorder, anorexia, bulimia, dementia, and learning disabilities. As depression, low mood, and anxiety

can potentially be more episodic conditions, further work was done to determine the timelines which should be used to define these cohorts. Based on local clinical and academic advice, and available evidence [11] we decided to create two patient cohorts for each conditions – patients who had ever received a diagnosis and patients who had a diagnosis recorded in the last 12 months.

Looking at the impact of the method of definition on activity in secondary care services indicated that the preferable way to define the depression, low mood, and anxiety cohorts was to include those who had ever had a diagnosis recorded in primary care. This was partly because only looking at those who had received a diagnosis in the past 12 months meant that we omitted a considerable proportion of patients who likely had common mental health disorders (CMDs, i.e. depression and/or anxiety) from our analyses. If required, it is possible to apply defined periods of time to conditions where full prolonged remission or cure is a possibility, e.g. cancer.

Prevalence rates for each condition in both Tower Hamlets and City & Hackney were calculated from the data. We assessed population prevalence in the dataset and compared with both the local and national population from other data sources. The prevalence rates of all 15 physical health conditions across six of the mental health conditions were also calculated in order to illustrate the bidirectional links between mental and physical health.

### *Health service utilisation*

Health service utilisation data outlined in this profile covered use of emergency services for exemplary purposes. Data relating to specific mental health conditions were described. Health service utilisation data for elective and non-elective inpatient admissions, and outpatient appointments are also available in the ELHCP Data Repository. These data come from the SUS dataset and provide number of attendances/admissions/appointments per patient per year. There is also data pertaining to associated spend (e.g. spend per activity, spend per patient per year) which is currently being validated and is therefore not presented in this paper.

### *Data completeness*

As the ELHCP Data Repository comprises routinely collected data from several different data flows there is considerable variation in completeness across variables. Completeness of certain variables appears to be better in patients with diagnoses of mental/physical health conditions compared to those with no diagnoses. We used type 2 diabetes and COPD as exemplar physical conditions, and depression and SMI as exemplar mental conditions to demonstrate how data

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

quality differs across patient groups. We provided completeness counts on several demographic and clinical variables including age, sex, deprivation level, ethnicity, smoking status, alcohol use, BMI, weight (kg), systolic blood pressure (SBP), diastolic blood pressure (DBP), and cholesterol. All variables were from primary care datasets. Alcohol use was measured using a composite variable of combined scores from the Alcohol Use Disorders Identification Test (AUDIT) and AUDIT-consumption (AUDIT-C) completed in primary care.

*Patient and public involvement and data security*

The primary purpose of this project was for the health and care sector to use for system planning, and effective commissioning, and therefore public stakeholders in this case comprises primary care managers, commissioners (GPs), public health experts and clinicians in primary and secondary care. A clinical expert group was convened consisting of fifteen clinicians, commissioners and public health experts from primary care and secondary care (mental health) services. This group met on a quarterly basis during initial profiling of the cohort and provided feedback into the analysis and emerging clinical and demographic characteristics.

The dataset contains sensitive patient level data; therefore, it was and remains important to ensure appropriate protections are in place. All data are de-identified, with no information such as name, home address or NHS number. Also, certain sensitive health data such as records relating to trauma and abuse, or specific sexual health data, are not visible.

To further preserve anonymity and to ensure proportionate and appropriate use of the data, access to records are limited to only being able to view aggregated records where there is a minimum of 5 individuals. Searches where there are fewer than 5 records are suppressed. Access to the complete de-identified records is further limited to specific individuals who are employed by, or under contract with, the health and care sector, have knowledge of how to use the data effectively and have undergone appropriate training.

This dataset has been compiled for the health and care sector to use for system planning. It contains sensitive patient level data; therefore, it is important to ensure appropriate protections are in place. Currently plans are being made to deploy the data into an electronic platform for internal use: recognising the possible value this data holds for researchers, there may be scope to support analysis deemed to be in the public interest, with appropriate data



protections in place. Currently, permissions must be sought on a case by case basis and approved by relevant IG governance groups.

### **Cohort description and findings to date**

#### *Population demographics*

The GP-registered populations for Tower Hamlets and City & Hackney in financial year 2017/2018 were 351,749 and 344,511, respectively. See Table 2 for a summary of population demographics for both Tower Hamlets and City & Hackney. Data for age, sex, and smoking status were recorded for the complete GP-registered populations in both East London boroughs. Deprivation index was recorded for most of the population (99.8% complete in both boroughs). Ethnicity data was recorded to a good level of completeness at 86.6% and 84.6% in Tower Hamlets and City & Hackney respectively. BMI data was more problematic and was recorded in the primary care records for less than a quarter of GP-registered patients (Tower Hamlets: 23.4%, City & Hackney: 24.0%).

Demographic data suggested that both Tower Hamlets and City & Hackney are 'young' boroughs relative to English national data[12] (median age 39 years) with mean ages of 32 (median = 31 years) and 34 years (median = 33 years) respectively. In terms of ethnicity, the majority of the GP-registered population in Tower Hamlets (38.2%) and City & Hackney (46.8%) were of white (or not stated) ethnicity. A considerable proportion of the population of Tower Hamlets were of Bangladeshi ethnicity (27.4%) and a significant amount of the City & Hackney population belonged to 'other ethnic group' (16.8%). Most of the GP-registered population in both East London boroughs were living in areas of high deprivation: in Tower Hamlets and City & Hackney 58.8% and 54.2% of the population respectively were in deciles 1 and 2, indicative of high deprivation. Approximately 20% of the population were recorded in primary care as being smokers. Although BMI is poorly recorded in primary care records, mean BMI scores indicated that the GP-registered population on average were overweight (Tower Hamlets: 26.3±8.1; City & Hackney: 27.1±8.6).

#### *Prevalence rates of mental and physical health conditions*

Prevalence rates for GP-coded diagnoses of mental health conditions in both Tower Hamlets and City & Hackney are provided in Figure 2. Prevalence rates were similar across both boroughs for all conditions. Comparative local and national prevalence data for each condition is provided in Table 3. The GP-coded prevalence of common mental health disorders (CMDs)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

was 16.8% and 17.5% in Tower Hamlets and City & Hackney respectively. In this case, CMD comprised diagnoses of depression, anxiety, as well as low mood. Low mood is often recorded both in place of and alongside depression in primary care and is therefore important to consider in a measure of CMD. For example, in Tower Hamlets of those with a diagnosis of depression 33% also had a diagnosis of low mood. CMD prevalence was in line with data from the latest Adult Psychiatric Morbidity Survey (APMS) which reported that 17% of the English population have a CMD[13]. However, the GP-coded prevalence of depression in Tower Hamlets (8.7%) and City & Hackney (11.4%) was higher than Public Health England (PHE) prevalence data[14] from the same areas from the same financial year (7.3%, 9.6%). This could be due to the GP-coded prevalence in this profile reflecting diagnoses of depression that were ever recorded, which likely led to somewhat inflated rates.

GP-coded prevalence of SMI (1.3%) from the ELHCP Data Repository was in line with data from the PHE Psychosis Data Report from 2016[9] which reported prevalence rates of SMI for both Tower Hamlets and City & Hackney at 1.32% and 1.36% respectively. Prevalence rates for learning disability were also in line with local and national data provided by Public Health England[14]. However, there were discrepancies between the GP-coded prevalence of personality disorder, eating disorders, and dementia in the ELHCP Data Repository and local and national PHE[14] and APMS[13] data. What this suggests is that these conditions might be under-recorded in primary care in these boroughs.

Prevalence rates for GP-coded diagnoses of physical health conditions in the ELHCP Data Repository are provided in Figure 3. Prevalence rates were similar across boroughs for most conditions. However, primary care records showed that levels of type 2 diabetes were higher in Tower Hamlets and levels of hypertension were higher in City & Hackney. Comparative local and national prevalence data for each condition is provided in Table 4. Local and national prevalence data was sourced from PHE public health profiles[14]. Overall, GP-coded prevalence of physical health conditions from the ELHCP Data Repository seems to be in line with PHE local data for the same areas. There is one notable discrepancy, however. Levels of asthma were much higher in the ELHCP Data Repository. This might be due to overdiagnosis of asthma in primary care which has been reported in both adults[15] and children[16]. Within the ELHCP Data Repository ‘liver disease’ comprises many liver complications, e.g. hepatitis, fatty liver disease, alcoholic liver disease. Therefore, it was not possible to directly compare prevalence data relating to liver disease with local and national statistics as PHE provide individual prevalence rates for each liver disease subtype.



### *Prevalence of physical health conditions among mental health conditions*

Mental and physical health are inextricably linked[17]. Having a mental health condition predisposes an individual to develop problems with their physical health. At the same time, living with a physical health condition can greatly impact mental health. Figure 4a and Figure 4b illustrate the prevalence of physical health conditions among mental health conditions in the ELHCP Data Repository, obtained from the N. Mental conditions included in this analysis were depression, low mood, anxiety, SMI, and personality disorder.

Results indicate that levels of all 15 physical health conditions were increased in those with mental health conditions. The distribution of the prevalence of physical health conditions amongst mental health conditions was similar across boroughs. The most pronounced increases in physical health conditions were seen for those flagged by their GP as having depression, SMI, and personality disorder. Some striking findings included the increase in levels of type 2 diabetes and hypertension among patients flagged by their GP as having SMI. Type 2 diabetes is known to be elevated in patients with SMI[18,19]. This increase might be brought about through use of antipsychotic medications known to be both diabetogenic and linked with weight gain[20], as well as through social deprivation, poor nutrition, chronic stress, and lifestyle changes[21]. Associations between SMI and hypertension reported in the literature have been more mixed[21]. Another striking finding that occurred in both boroughs was the high prevalence of asthma for people with a primary care diagnosis of personality disorder. This may be linked to smoking which is known to be increased in patients with personality disorder[22]. Personality disorder has been linked to hypochondriasis and perceived breathlessness[23], which might also be contributing factors to higher levels of asthma diagnosis in primary care for these patients.

### *A&E attendances across mental health conditions*

We used A&E as an example of a health service utilisation outcome measure to illustrate some of the activity data available in the ELHCP Data Repository. Data relating to use of A&E across both boroughs are detailed in Table 5. Based on crude attendance rates (per 1000 population), patients with mental health conditions attended A&E 54% and 76% more in Tower Hamlets and City & Hackney respectively compared to those without GP-recorded mental health conditions. Increased attendance rates were particularly pronounced in patients with dementia, personality disorder, and SMI. Similar health service utilisation data exists relating to inpatient admissions and use of outpatient services.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

*Data completeness*

The ELHCP Data Repository relies on data inputted by health care provider organisations (GP practices, clinical coders in acute care settings, etc.). It is comprised of routinely collected data that can vary in completeness across variables. Examination of four example physical and mental health conditions illustrated how data completeness varies across patient groups. Table 6 details completeness on key demographic and clinical variables for patients with type 2 diabetes, COPD, depression, and SMI in Tower Hamlets. These data were all recorded in primary care databases. Data pertaining to age, sex, deprivation level, ethnicity, and smoking status were complete for all patients included in the analysis. Data relating to alcohol use and clinical factors differed between those with and without health conditions in that the data was considerably more complete for patients registered as having a condition. This improvement in completeness is likely down to the implementation of financial incentive schemes introduced as part of the UK’s Quality Outcomes Framework (QOF) to improve clinical management of patients with chronic conditions[4,24]. What this illustrates is that the ELHCP Data Repository is well-suited to perform research on individual physical and mental health conditions using primary care data, where data completeness is to a high standard.

**Strengths and limitations**

*Strengths*

One of the main strengths of the dataset is that it provides the opportunity to analyse a relatively young, urban, ethnically diverse, and considerably deprived population: groups that are hard to reach with traditional survey methods. Given patient records are linked across care settings, the data can be used to evaluate the impacts of integrated interventions. In this profile we have illustrated that the ELHCP Data Repository is representative in terms of the prevalence of most mental and physical health conditions compared to other sources of local and national prevalence data supporting its suitability for real-world research and decision-making. This dataset also allows for the analysis of comorbidities and multimorbidity. Moreover, there is a large amount of data available around health service utilisation and associated spend meaning that the dataset can be used to understand what drives use and spend and support healthcare commissioning and health service delivery.

The dataset is in regular use by commissioners to better understand the health and care needs of the population, highlight where there is scope for improved patient care, and evaluate the impact of health and care interventions of patients and the system. This has provided

commissioners, and others working on health and care system integration, with more sophisticated evidence and analysis to inform decisions about where to target improvement efforts to maximise benefit to both patients and the system. The findings presented in this cohort profile have highlighted where links between mental and physical health are most prevalent in the population and helped to shape commissioning priorities accordingly. For example, analysis highlighting mental and physical comorbidity supported conversations about expanding IAPT services within long term condition pathways, noting that it may be particularly impactful for people with type 2 diabetes, respiratory conditions, and cancer. Furthermore, over the next year Tower Hamlets is investing in a new 24/7 core fidelity crisis service, and the transformation of community mental health services to become embedded in primary care networks. These new models will support the assessment and treatment of patients at the right place at the right time. The ELHCP Data Repository described in this cohort profile will be an important tool in evaluating the impact these service changes have on patients and the healthcare system.

In terms of completeness, this paper highlighted that data were considerably more complete for patients with long-term mental or physical illness (i.e. type 2 diabetes, COPD, depression, SMI) compared to patients without. This supports the suitability of the dataset for the analysis of patients with long-term conditions.

### *Limitations*

In terms of limitations associated with the dataset, in some cases prevalence data was not in line with local and national data, meaning that it is possible some conditions are under-recorded in primary care, i.e. personality disorder, eating disorders, and dementia. Also, as prevalence data relies on GP diagnoses, it is not possible to gauge disease duration in patients as the date in which a diagnosis is first recorded might not be the date the patient received the diagnosis. For example, the diagnosis might have been recorded on the day the patient registered with their GP, reflecting that date rather than the date of diagnosis. In terms of data completeness, although this is good for patients with long-term conditions, there is considerable missingness among 'healthy' patients in the dataset. This has also been reported in a study of a large primary care databases in the UK which partially attributed the discrepancy in missing data to GP incentive schemes [25]. It is possible to resolve this with imputation to some degree. As is the case with all routine clinical data, the dataset requires extensive cleaning to remove duplicate entries and spurious values out of clinical range (e.g. BMI, blood pressure) prior to analysis.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Another limitation involves the amount of data available. Currently, there are two financial years of data that are at an acceptable standard for research and/or decision support. New data continues to be collected and flow into this data set, so the time series available will increase over time. A further limitation surrounds the ability to measure population mobility in a dataset of this kind. One way is to look at the length of GP registration as a proxy for this, but this is problematic as it is not possible to determine whether the mobility is within the borough, or whether patients have come from outside.

**Conclusions and future collaborations**

This cohort profile describes a unique patient-linked dataset in East London which can be used to facilitate the development of similar datasets in other parts of the UK. Moreover, this dataset provides a unique opportunity to analyse a relatively young, urban, diverse, and deprived population in East London. The dataset is growing year on year and data quality should improve over time making this an invaluable resource for researchers, clinicians, commissioners, and policy makers. The data are being further cleaned and evaluated using imputation, Bayesian, and economic methods, including engagement with machine learning work funded by the Alan Turing Institute to improve and automate the analysis of large scale electronic medical records (KNIFE project).

**Author statement**

AR drafted the manuscript with input and critical revisions from all authors. QK, KB, EW, GH, RF, MF, and KB provided supervision throughout. KB is the principal investigator and grant holder. Analysis of the data was performed by AA, EC, and IE. All authors reviewed the paper before submission.

**References**

- 1 Cook JA, Collins GS. The rise of big clinical databases. *Br J Surg* 2015;**102**:e93–101. doi:10.1002/bjs.9723
- 2 Bates DW, Saria S, Ohno-Machado L, *et al.* Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Aff Proj Hope* 2014;**33**:1123–31. doi:10.1377/hlthaff.2014.0041
- 3 Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst* 2014;**2**:3. doi:10.1186/2047-2501-2-3

- 4 Correa A, Hinton W, McGovern A, *et al*. Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) sentinel network: a cohort profile. *BMJ Open* 2016;**6**:e011092. doi:10.1136/bmjopen-2016-011092
- 5 Herrett E, Gallagher AM, Bhaskaran K, *et al*. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015;**44**:827–36. doi:10.1093/ije/dyv098
- 6 Perera G, Broadbent M, Callard F, *et al*. Cohort profile of the South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLaM BRC) Case Register: current status and recent enhancement of an Electronic Mental Health Record-derived data resource. *BMJ Open* 2016;**6**:e008721. doi:10.1136/bmjopen-2015-008721
- 7 Tower Hamlets Council. Ethnicity in Tower Hamlets: Analysis of 2011 census data. 2013. [https://www.towerhamlets.gov.uk/Documents/Borough\\_statistics/Ward\\_profiles/Census-2011/RB-Census2011-Ethnicity-2013-01.pdf](https://www.towerhamlets.gov.uk/Documents/Borough_statistics/Ward_profiles/Census-2011/RB-Census2011-Ethnicity-2013-01.pdf) (accessed 7 Mar 2019).
- 8 Trust for London. Poverty and inequality data for Tower Hamlets. Trust Lond. 2017./data/boroughs/tower-hamlets-poverty-and-inequality-indicators/ (accessed 3 Jul 2019).
- 9 Public Health England. Psychosis data report. 2016;:77.
- 10 Noble M, McLennan D, Wilkinson K, *et al*. English Indices of Deprivation. London: : Communities and Local Government 2007.
- 11 John A, McGregor J, Fone D, *et al*. Case-finding for common mental disorders of anxiety and depression in primary care: an external validation of routinely collected data. *BMC Med Inform Decis Mak* 2016;**16**:35. doi:10.1186/s12911-016-0274-7
- 12 Ethnicity facts and figures: Age groups. HM Government Services 2018. <https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/demographics/age-groups/latest#download-the-data> (accessed 6 Jun 2019).
- 13 Adult Psychiatric Morbidity Survey: Survey of Mental Health and Wellbeing, England, 2014. NHS Digit. <https://digital.nhs.uk/data-and-information/publications/statistical/adult-psychiatric-morbidity-survey/adult-psychiatric-morbidity-survey-survey-of-mental-health-and-wellbeing-england-2014> (accessed 11 Jan 2019).
- 14 Public Health England. Public Health Profiles. <https://fingertips.phe.org.uk/profile-group/mental-health> (accessed 18 Jun 2019).
- 15 Aaron SD, Vandemheen KL, Boulet L-P, *et al*. Overdiagnosis of asthma in obese and nonobese adults. *CMAJ* 2008;**179**:1121–31. doi:10.1503/cmaj.081332
- 16 Akker IL den, Luijn K van, Verheij T. Overdiagnosis of asthma in children in primary care: a retrospective analysis. *Br J Gen Pract* 2016;**66**:e152–7. doi:10.3399/bjgp16X683965
- 17 Doherty AM, Gaughran F. The interface of physical and mental health. *Soc Psychiatry Psychiatr Epidemiol* 2014;**49**:673–82. doi:10.1007/s00127-014-0847-7

18 Das-Munshi J, Ashworth M, Gaughran F, *et al.* Ethnicity and cardiovascular health inequalities in people with severe mental illnesses: protocol for the E-CHASM study. *Soc Psychiatry Psychiatr Epidemiol* 2016;**51**:627–38. doi:10.1007/s00127-016-1185-8

19 Ward M, Druss B. The epidemiology of diabetes in psychotic disorders. *Lancet Psychiatry* 2015;**2**:431–51. doi:10.1016/S2215-0366(15)00007-3

20 Gianfrancesco FD, Grogg AL, Mahmoud RA, *et al.* Differential effects of risperidone, olanzapine, clozapine, and conventional antipsychotics on type 2 diabetes: findings from a large health plan database. *J Clin Psychiatry* 2002;**63**:920–30.

21 Osborn DPJ, Wright CA, Levy G, *et al.* Relative risk of diabetes, dyslipidaemia, hypertension and the metabolic syndrome in people with severe mental illnesses: systematic review and metaanalysis. *BMC Psychiatry* 2008;**8**:84. doi:10.1186/1471-244X-8-84

22 Zvolensky MJ, Jenkins EF, Johnson KA, *et al.* Personality Disorders and Cigarette Smoking among Adults in the United States. *J Psychiatr Res* 2011;**45**:835–41. doi:10.1016/j.jpsychires.2010.11.009

23 Chetta A, Gerra G, Foresi A, *et al.* Personality Profiles and Breathlessness Perception in Outpatients with Different Gradings of Asthma. *Am J Respir Crit Care Med* 1998;**157**:116–22. doi:10.1164/ajrccm.157.1.9702093

24 Kontopantelis E, Reeves D, Valderas JM, *et al.* Recorded quality of primary care for patients with diabetes in England before and after the introduction of a financial incentive scheme: a longitudinal observational study. *BMJ Qual Saf* 2013;**22**:53–64. doi:10.1136/bmjqs-2012-001033

25 Petersen I, Welch CA, Nazareth I, *et al.* Health indicator recording in UK primary care electronic health records: key implications for handling missing data. *Clin Epidemiol* 2019;**11**:157–67. doi:10.2147/CLEP.S191437

**Table 1.** Mental and physical health conditions recorded in primary care records**Mental health conditions**

Depression; low mood\*; anxiety; severe mental illness (SMI); personality disorder; anorexia; bulimia; dementia; learning disability

**Physical health conditions**

Type 1 diabetes, type 2 diabetes, asthma, chronic obstructive pulmonary disorder (COPD), cancer (excluding basal cell carcinoma), stroke, coronary heart disease (CHD), heart failure, atrial fibrillation, hypertension, peripheral artery disease (PAD), chronic kidney disease (CKD), liver/pancreas complications (e.g. hepatitis, liver disease, chronic pancreatitis), osteoporosis, epilepsy

\*low mood is often recorded in primary care in place of depression and is therefore important to consider in any analyses relating to mental health



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Table 2.** Population demographics for the GP-registered populations of Tower Hamlets and City & Hackney

	Tower Hamlets	City & Hackney
<b>GP registered population (n,%)</b>	351749 (100%)	344511 (100%)
<b>Age (years) (M±SD)</b>	32 ± 17.97	34 ± 19.47
<b>Sex (n,%)</b>		
<i>Male</i>	178088 (50.63%)	168371 (48.87%)
<i>Female</i>	173656 (49.37%)	176134 (51.13%)
<b>Ethnicity (n,%)</b>		
<i>White or not stated</i>	134411 (38.21%)	161069 (46.75%)
<i>Indian</i>	9295 (2.64%)	9067 (2.63%)
<i>Pakistani</i>	3547 (1.01%)	2707 (0.79%)
<i>Bangladeshi</i>	96418 (27.41%)	5877 (1.71%)
<i>Other Asian</i>	8151 (2.32%)	6189 (1.80%)
<i>Black Caribbean</i>	4281 (1.22%)	16052 (4.66%)
<i>Black African</i>	14034 (3.99%)	30037 (8.72%)
<i>Chinese</i>	10435 (2.97%)	4137 (1.20%)
<i>Other ethnic group</i>	23934 (6.80%)	56140 (16.29%)
<b>Deprivation index (n,%)</b>		
<i>High deprivation*</i>	206723 (58.77%)	186784 (54.22%)
<i>Low deprivation**</i>	4955 (1.41%)	3134 (0.91%)
<b>Smokers (n,%)</b>	69627 (19.79%)	72371 (21.01%)
<b>BMI (kg/m<sup>2</sup>) (M±SD)</b>	26.34 ± 8.28	27.06 ± 8.58

\*High deprivation refers to the GP-registered population in decile 1 and 2  
\*\*Low deprivation refers to the GP-registered population in decile 9 and 10  
BMI = body mass index; M: Mean; SD: Standard deviation



Table 3. GP-coded prevalence of mental health conditions in Tower Hamlets and City & Hackney compared to local and national prevalence data (sources: APMS 2014, PHE Public Health Profiles, PHE Psychosis Data Report 2016)

Condition	GP-coded prevalence				Local prevalence		National prevalence
	<i>Tower Hamlets</i>	<i>95% CI</i>	<i>City &amp; Hackney</i>	<i>95% CI</i>	<i>Tower Hamlets</i>	<i>City &amp; Hackney</i>	<i>England</i>
<b>CMDs</b>	16.8%	16.69,16.94	17.5%	17.37,17.62	15.4%[14]	17%[14]	17%[13]
<b>SMI</b>	1.3%	1.24,1.31	1.3%	1.31,1.38	1.32%[9]	1.36%[9]	0.5-1.5%[9]
<b>Personality disorder</b>	0.5%	0.49,0.53	0.6%	0.55,0.60	-	-	13.7%[13]*
<b>Eating disorders</b>	0.3%	0.29,0.33	0.4%	0.37,0.41	6.9%[14]	7.4%[14]	6.9%[14]
<b>Dementia</b>	0.4%	0.34,0.38	0.4%	0.38,0.42	5.25%[14]	4.37%[14]	4.33%[14]
<b>Learning disability</b>	0.3%	0.32,0.36	0.4%	0.36,0.41	0.3%[14]	0.4%[14]	0.5%[14]

\*Percentage of sample who scored  $\geq 4$  on the Standardised Assessment of Personality – Abbreviated Scale (SAPAS)

APMS = Adult Psychiatric Morbidity Survey; CMDs = common mental health disorders; GP = general practitioner; PHE = Public Health England; SMI = severe mental illness

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46

**Table 4.** GP-coded prevalence of physical health conditions in Tower Hamlets and City & Hackney compared to local and national prevalence data (source: PHE Public Health Profiles)

Condition	GP-coded prevalence		Local prevalence		National prevalence
	<i>Tower Hamlets</i>	<i>City &amp; Hackney</i>	<i>Tower Hamlets</i>	<i>City &amp; Hackney</i>	<i>England</i>
Type 1 diabetes*	0.3%	0.4%			
Type 2 diabetes*	5.1%	4.4%	6.7%	5.7%	6.8%
Asthma	8.7%	9.2%	4.3%	4.4%	5.9%
COPD	1.3%	1.2%	1.2%	1.1%	1.9%
Cancer	1.4%	1.8%	1.1%	1.4%	2.7%
Stroke	0.8%	0.9%	0.9%	0.7%	1.8%
CHD	1.6%	1.7%	1.5%	1.5%	3.1%
Heart failure	0.7%	0.9%	0.4%	0.7%	0.8%
Atrial fibrillation	0.6%	0.8%	0.6%	0.7%	1.9%
Hypertension	7.4%	9.3%	7.5%	9.3%	13.9%
PAD**	0.3%	0.4%	0.96%	1.01%	1.16%
CKD	2.0%	1.8%	2.5%	2.1%	4.1%
Liver disease†	1.9%	1.4%	-	-	-
Osteoporosis	0.06%	0.09%	-	-	0.6%
Epilepsy	0.8%	0.8%	0.4%	0.5%	0.8%

\*PHE does not provide prevalence rates for diabetes subtypes. Therefore local and national prevalence rates is for diabetes overall.

\*\*Local and national prevalence data for PAD from PHE were based on modelled estimated prevalence

†It was not possible to provide local and national prevalence rates for liver disease as data is provided separately by PHE for disease subtype (e.g. hepatitis C, fatty liver disease, alcoholic liver disease)

CHD = coronary heart disease; CKD = chronic kidney disease; COPD = chronic obstructive pulmonary disorder; GP = general practitioner; PAD = peripheral artery disease; PHE = Public Health England

For peer review only

**Table 5.** The impact of mental health conditions on A&E attendances, financial year 2017/2018

Accident & Emergency								
Activity (attendances)								
	Total attendances		% Share		Crude rate (per 1000 pop.)			
	TH	C&H	TH	C&H	TH	95% CI	C&H	95% CI
Population	159201	161641	100%	100%	452	412, 495	469	428, 513
All non-mental health	118272	116816	74.3%	72.3%	412	373, 453	413	375, 454
All mental health	40929	44825	25.7%	27.7%	635	587, 686	728	677, 782
CMDs	31625	34215	19.9%	21.2%	535	491, 582	568	523, 616
SMI	3800	4383	2.4%	2.7%	847	791, 906	945	886, 1007
Personality disorder	2529	2914	1.6%	1.8%	1413	1341, 1488	1467	1393, 1543
Eating disorder	584	549	0.4%	0.3%	554	509, 602	424	385, 466
Dementia	1595	1471	1.0%	0.9%	1248	1180, 1319	1075	1012, 1141
Learning disability	796	1293	0.5%	0.8%	670	621, 722	974	914, 1037

A&E = Accident & Emergency; CMDs = common mental health disorders; C&H = City & Hackney; SMI = severe mental illness; TH = Tower Hamlets

**Table 6.** ELHLP data completeness (%) across physical and mental health conditions in Tower Hamlets, financial year 2017/2018

	<b>T2DM</b>	<b>No T2DM</b>	<b>COPD</b>	<b>No COPD</b>	<b>Depression</b>	<b>No Depression</b>	<b>SMI</b>	<b>No SMI</b>
<i>Variables</i>	<i>% complete</i>	<i>% complete</i>	<i>% complete</i>	<i>% complete</i>	<i>% complete</i>	<i>% complete</i>	<i>% complete</i>	<i>% complete</i>
<b>Age</b>	100%	100%	100%	100%	100%	100%	100%	100%
<b>Sex</b>	100%	100%	100%	100%	100%	100%	100%	100%
<b>Deprivation</b>	100%	100%	100%	100%	100%	100%	100%	100%
<b>Ethnicity</b>	100%	100%	100%	100%	100%	100%	100%	100%
<b>Smoking status</b>	100%	100%	100%	100%	100%	100%	100%	100%
<b>Alcohol use</b>	94.6%	51.1%	91.8%	52.8%	73.1%	51.4%	89.2%	52.8%
<b>BMI</b>	99.3%	72.6%	99.3%	73.7%	95.4%	71.9%	97.5%	73.7%
<b>Weight</b>	99.4%	84.6%	99.5%	85.2%	96.7%	84.3%	98.2%	85.2%
<b>SBP</b>	99.6%	68.8%	99.6%	69.9%	96.4%	67.9%	98.5%	69.9%
<b>DBP</b>	99.6%	68.7%	99.6%	69.9%	96.3%	67.9%	98.5%	69.9%
<b>Cholesterol</b>	99.4%	28.8%	96.9%	31.6%	64.9%	29.3%	84.6%	37.8%

BMI = body mass index; COPD: chronic obstructive pulmonary disorder; DBP = diastolic blood pressure; ELHLP = East London Healthcare Links Project; SBP = systolic blood pressure; SMI = severe mental illness; T2DM = Type 2 diabetes mellitus

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Figure 1.** Diagram of data flows from healthcare providers to Tower Hamlets & City Hackney CCGs

CCG = Clinical Commissioning Group, CDS = Commissioning Data Sets, CSU = Commissioning Support Unit, DCB = Data Coordination Board, DSCRO = Data Services for Commissioners Regional Office, NEL = North East London, NHS = National Health Service

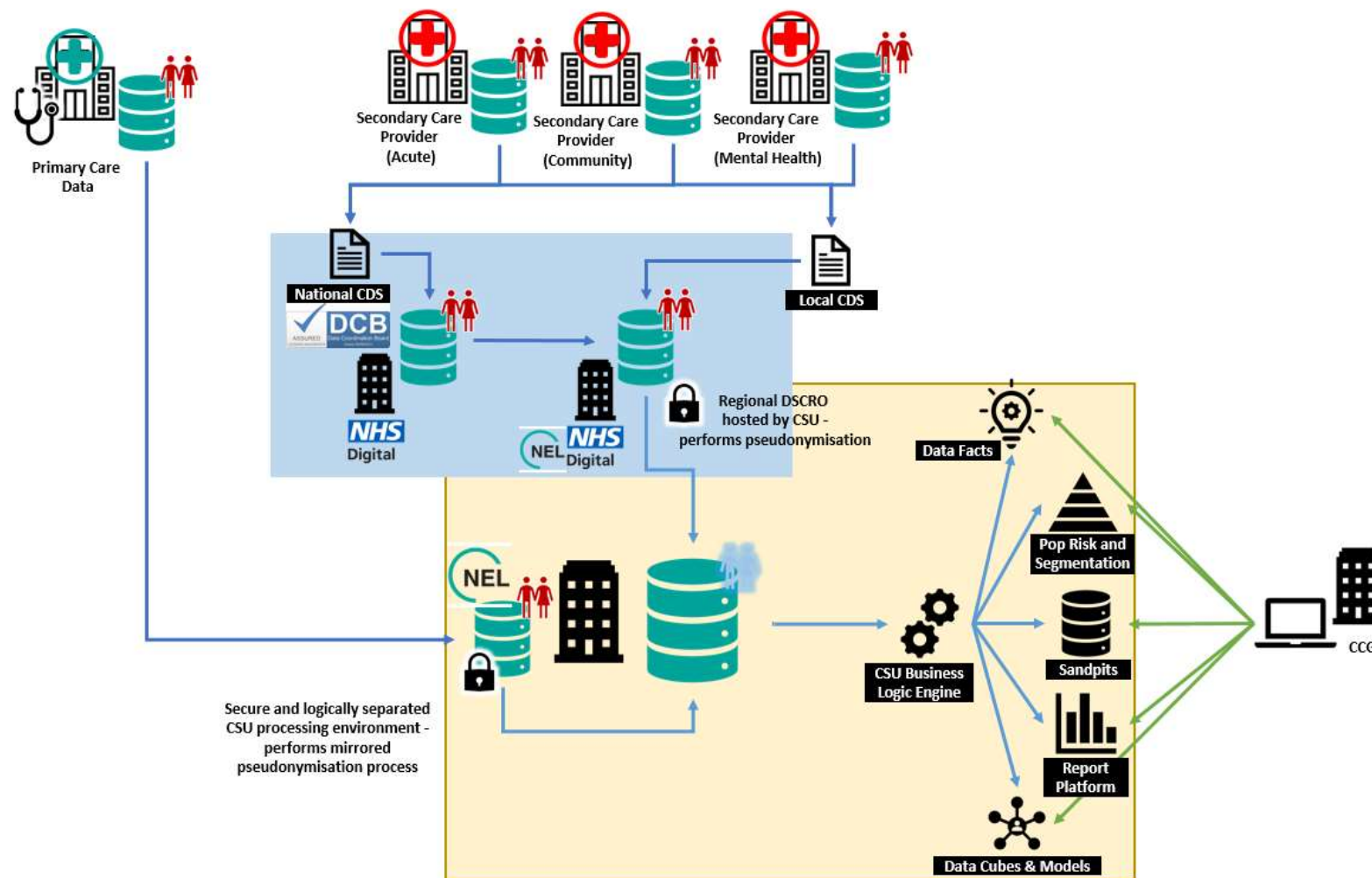
**Figure 2.** GP-coded prevalence rates of mental health conditions in Tower Hamlets and City & Hackney, financial year 2017/2018.

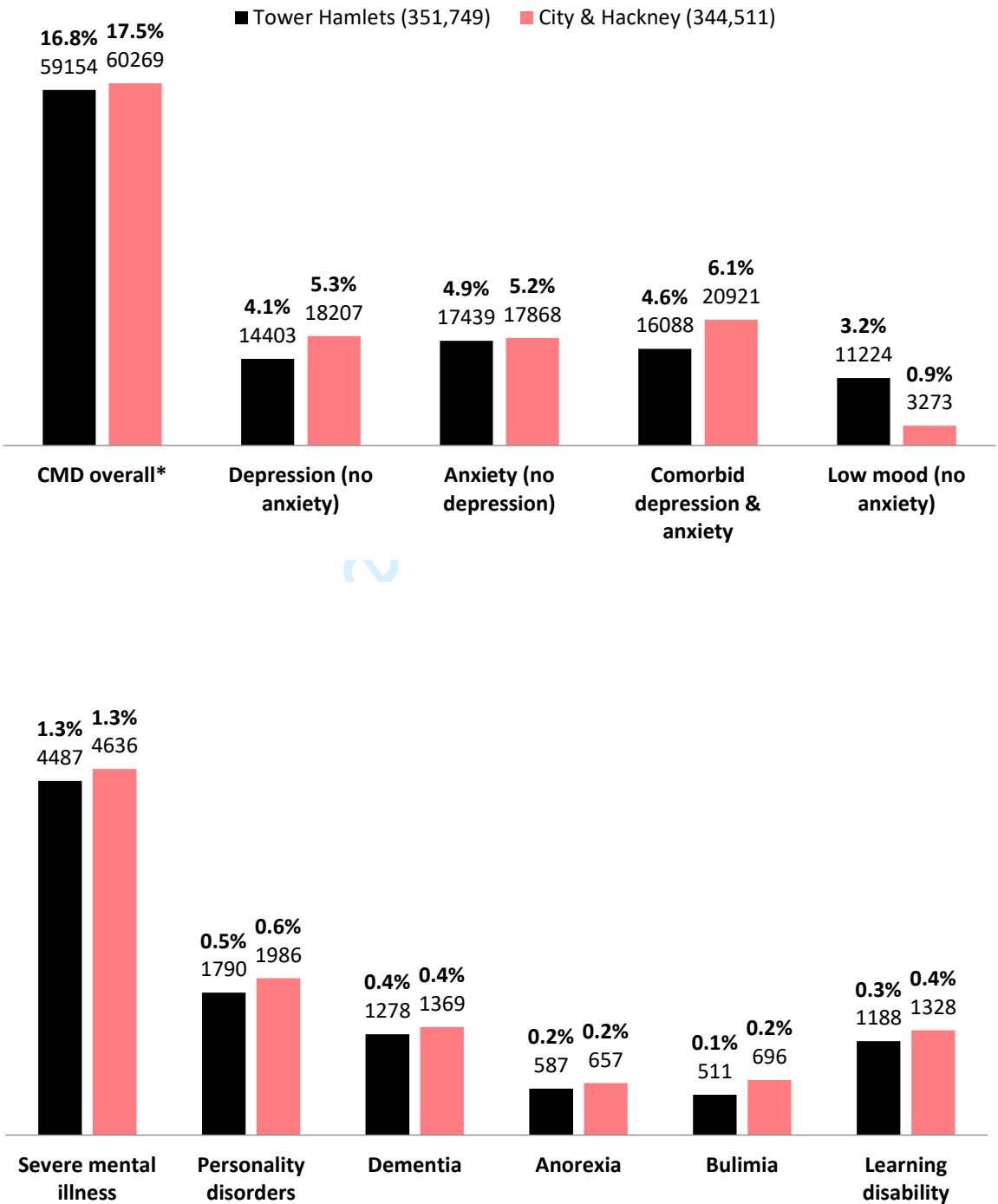
\*CMD incorporates depression, anxiety and low mood

**Figure 3.** GP-coded prevalence rates of physical health conditions in Tower Hamlets and City & Hackney, financial year 2017/2018

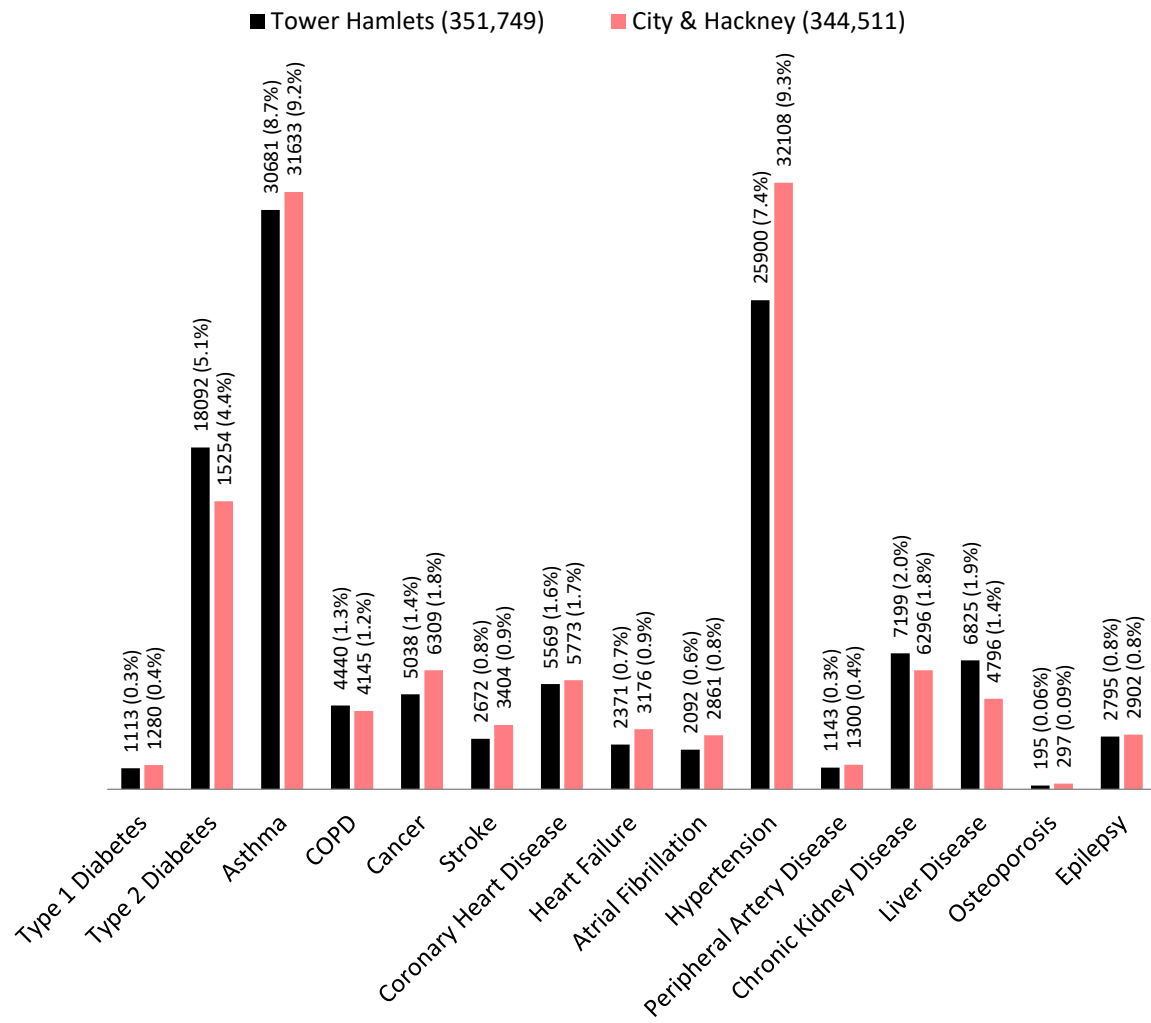
**Figure 4a.** Prevalence of physical health conditions among mental health conditions in Tower Hamlets

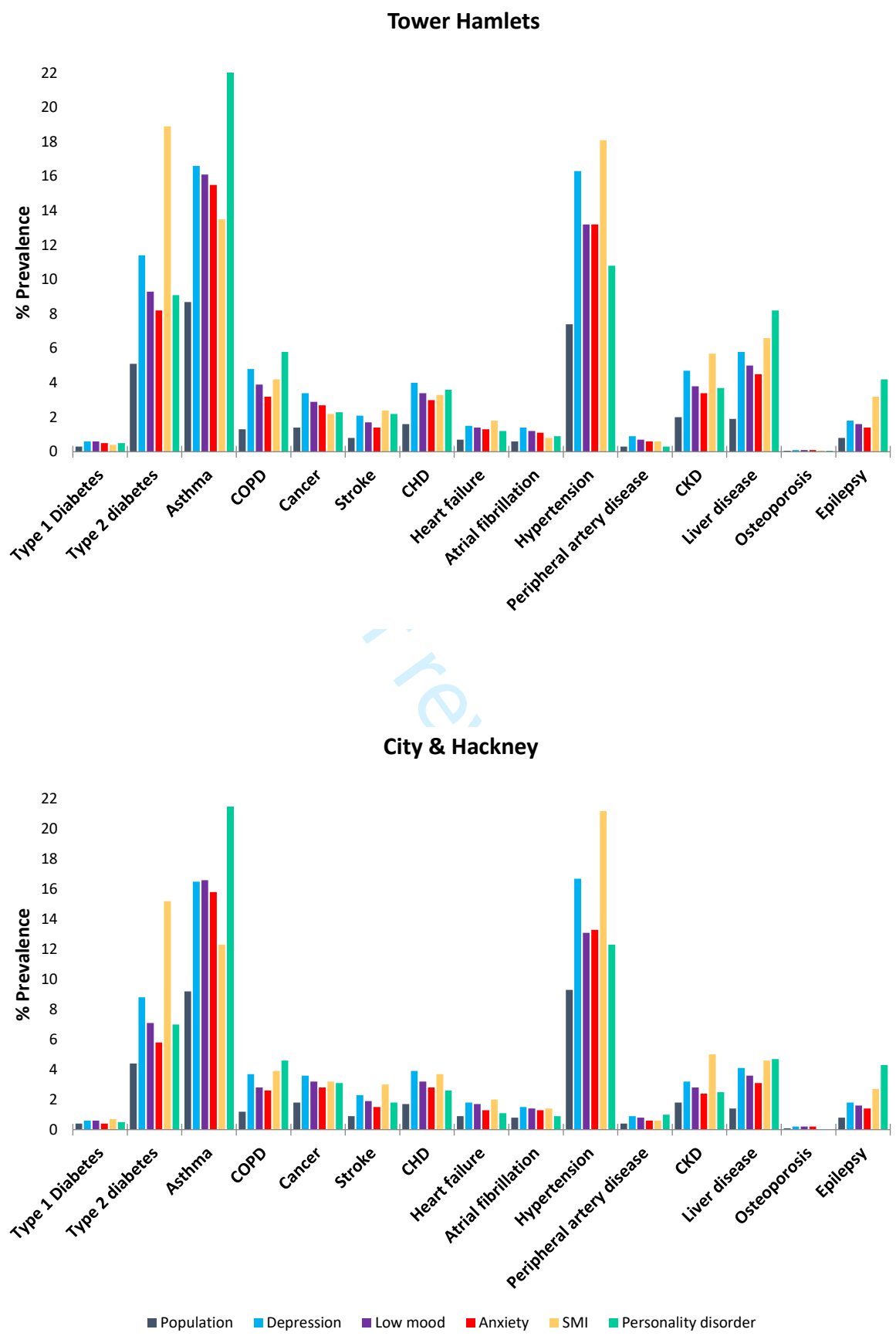
**Figure 4b.** Prevalence of physical health conditions among mental health conditions in City & Hackney











# BMJ Open

## Cohort profile: The East London Health and Care Partnership Data Repository - Using novel integrated data to support commissioning and research

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2020-037183.R2
Article Type:	Cohort profile
Date Submitted by the Author:	04-Aug-2020
Complete List of Authors:	Ronaldson, Amy; King's College London, Institute of Psychiatry, Psychology and Neuroscience; Queen Mary University of London, Centre for Psychiatry Chandakas, Evangelos; NHS Tower Hamlets Clinical Commissioning Group Kang, Qiongwen; NHS Tower Hamlets Clinical Commissioning Group Brennan, Katie; NHS Tower Hamlets Clinical Commissioning Group Akande, Aminat; NHS Tower Hamlets Clinical Commissioning Group Ebyarimpa, Irene; NHS Tower Hamlets Clinical Commissioning Group Wyllie, Eleanor; Healthy London Partnership Howard, George; Healthy London Partnership Fradgley, Richard; East London NHS Foundation Trust Freestone, Mark; Barts and The London School of Medicine & Dentistry Queen Mary University of London, Centre for Psychiatry at the Wolfson Institute of Preventive Medicine Bhui, Kamaldeep; Barts and The London School of Medicine & Dentistry Queen Mary University of London, Centre for Psychiatry at the Wolfson Institute of Preventive Medicine
<b>Primary Subject Heading</b>:	Health informatics
Secondary Subject Heading:	Health services research, Health policy, Public health, Health economics
Keywords:	Health economics < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Health policy < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, MENTAL HEALTH, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, EPIDEMIOLOGY, PRIMARY CARE

SCHOLARONE™  
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

## **Cohort profile: The East London Health and Care Partnership Data Repository - Using novel integrated data to support commissioning and research**

Amy Ronaldson, PhD<sup>1,6\*</sup>, Evangelos Chandakas, PhD<sup>2</sup>, Qiongwen Kang, MSc<sup>2</sup>, Katie Brennan, MSc<sup>2</sup>, Aminat Akande<sup>2</sup>, Irene Ebyarimpa<sup>2</sup>, Eleanor Wyllie<sup>3</sup>, George Howard<sup>3</sup>, Richard Fradgley<sup>4</sup>, Mark Freestone, PhD<sup>1,4,5</sup>, Kamaldeep Bhui, MD<sup>1</sup>

<sup>1</sup>Centre for Psychiatry, Wolfson Institute of Preventive Medicine, Barts & The London School of Medicine, Old Anatomy Building, Charterhouse Square, London EC1M 6BQ

<sup>2</sup>Tower Hamlets Clinical Commissioning Group, Mile End Hospital, Bancroft Road, Bethnal Green, London E1 4DG

<sup>3</sup>Healthy London Partnership, 75-77 Worship Street, Hackney, London EC2A 2DU

<sup>4</sup>East London NHS Foundation Trust, Trust Headquarters, Robert Dolan House, 9 Alie Street, London E1 8DE

<sup>5</sup>The Alan Turing Institute, British Library, London NW1 2DB

<sup>6</sup>Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, De Crespigny Park, London SE5 8AF

**Keywords:** Mental health; health services; research cohort; commissioning; health policy

**Word count:** 3,505

**Number of tables and figures:** 10

**\*Corresponding author and person to whom reprint requests should be addressed:**

Dr Amy Ronaldson, Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, De Crespigny Park, London SE5 8AF

Email: [amy.ronaldson@kcl.ac.uk](mailto:amy.ronaldson@kcl.ac.uk)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Abstract**

**Purpose:** The East London Health and Care Partnership (ELHCP) Data Repository was established to support commissioning decisions in London. This dataset comprises routine clinical data for the GP-registered populations of two London boroughs, Tower Hamlets and City & Hackney, and provides a rich source of demographic, clinical, and health service use data of relevance to clinicians, commissioners, researchers and policy makers. This paper describes the dataset in its current form, its representativeness, and data completeness.

**Participants:** There were 351,749 and 344,511 members of the GP-registered population in the two boroughs respectively for the financial year 2017/18. Demographic information and prevalence data were available for nine mental health and 15 physical health conditions. Prevalence rates from the cohort were compared with local and national data. In order to illustrate the health service use data available in the dataset, emergency department use across mental health conditions was described. Information about data completeness was provided.

**Findings to date:** The ELHCP Data Repository provides a rich source of information about a relatively young, urban, ethnically diverse, population within areas of socioeconomic deprivation. Prevalence data were in line with local and national statistics with some exceptions. Physical health conditions were more common in those with mental health conditions, reflecting that co-morbidities are the norm rather than the exception. This has implications for integrated care. Data completeness for risk factors (e.g. blood pressure, cholesterol) was high in patients with long-term conditions.

**Future plans:** The data are being further cleaned and evaluated using imputation, Bayesian, and economic methods, principally focusing on specific cohorts, including type II diabetes, depression and personality disorder. Data continue to be collected for the foreseeable future to support commissioning decisions which will also enable more long-term prospective analysis as data become available at the end of each financial year.

### **Strengths and limitations of this study**

- This dataset will allow for the analysis of a relatively young, urban, ethnically diverse, considerably deprived population – groups that are hard to research with traditional survey methods.
- The data are linked across care settings, allowing for the evaluation of integrated interventions.
- Large amounts of data around health service utilisation and associated spend means this dataset can be used to support healthcare commissioning and health service delivery.
- Reliance on primary care data for diagnostic information means that prevalence rates can be under-estimated and certain information can be missed.
- As is the case with all routine clinical data, the dataset requires extensive cleaning to remove errors.

### **Introduction**

The use of observational data is fundamental to healthcare service provision and health research, and increasingly large amounts of routinely collected clinical data are becoming available. The main advantage of these datasets are their comprehensive nature, and large patient numbers [1]. The development of these large routine datasets also allow for the application of big data analytic techniques within health research [2]. These techniques can evaluate population characteristics, identify risk factors, and allow for the development of predictive models relating to clinical outcomes [1,3]. Moreover, these datasets identify areas of high health service utilisation and high-cost patients which might support decision-making around health service delivery for commissioners, policy makers, and healthcare providers. To date, several large datasets comprising routine clinical data have been developed in the United Kingdom (UK); these are being used to inform clinical practice and healthcare delivery [4–6].

Recently, a large-scale routine clinical dataset has been developed in the Tower Hamlets borough of London, UK, comprising data from the general practitioner (GP)-registered populations of Tower Hamlets and its neighbouring borough, City & Hackney. The East London Health and Care Partnership (ELHCP) Data Repository was developed primarily to support commissioning decisions and health service delivery in East London and comprises rich information relevant to population health and clinical research. These areas of London are unique inner-city areas with diverse populations and high levels of deprivation. Tower Hamlets is home to the largest Bangladeshi community in England [7]. It also has the highest rate of

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

poverty, child poverty, and unemployment of any London borough [8]. What might be of interest to researchers of severe mental illness is that the borough of City & Hackney has the highest incidence of psychosis in England [9].

The aim of this cohort profile is to introduce and describe the dataset in Tower Hamlets and City & Hackney, comparing population demographics, prevalence rates of mental and physical health conditions, health service utilisation, and data completeness. We believe this cohort profile might facilitate the development of other similar datasets. Moreover, these data will prove useful for clinicians, researchers, commissioners, and policy makers, particularly those with an interest in understanding health in urban, diverse, and deprived areas.

*The development of the ELHCP Data Repository*

The ELHCP Data Repository was developed by Tower Hamlets Clinical Commissioning Group (CCG) (<https://www.towerhamletsccg.nhs.uk/>) and North East London Commissioning Support Unit (NEL CSU) (<http://www.nelcsu.nhs.uk/>), with funding provided by Tower Hamlets CCG and Tower Hamlets Together (THT) Vanguard (<https://www.towerhamletstogether.com/>). Development of the dataset started in 2015 and efforts to expand the dataset (e.g. to all areas in North East London) and improve data quality are ongoing.

The dataset comprises up-to-date patient level data for GP-registered populations in the London boroughs of Tower Hamlets and City & Hackney and is updated every three months. All patient data is depersonalised and linked by replacing patient NHS numbers with unique ID numbers. Data pertaining to geographic area (i.e. Lower Super Output Area (LSOA)) and other population-level data are drawn from data aggregated by the Office of National Statistics (ONS). Presently, secure data streams are in place for: acute (inpatient, outpatient, and emergency) services; community services; mental health services; substance misuse services; and primary care services (e.g. diagnostic information, prescribing). Efforts are underway to establish secure data streams for social care data, public health data (e.g. sexual health), and urgent care data (e.g. walk-in-centres). However, challenges surrounding linking social care data means that sensitive information relating to housing and benefits will not be integrated into the dataset.

As illustrated in Figure 1, data for acute, community, and mental health services flow to the CCG via NHS digital, The Data Services for Commissioners Regional Office (DSCRO), and



local CSUs. Pseudonymisation of the data is performed by the DSCRO hosted by the CSU. The data then flows to the local CSU which formats and organises the data for use at the CCG. Primary care data flows to NEL CSU where a pseudonymisation process is carried out which mirrors that performed by the DSCRO. These data then flow to the local CSU alongside the acute, community, and mental health data where they are added to the ELHCP Data Repository.

## **Methods**

This cohort profile is a descriptive piece detailing the following:

- (i) Demographic information (e.g. age, sex, ethnicity, and deprivation)
- (ii) Prevalence rates of mental and physical health conditions
- (iii) Data relating to health service utilisation in patients with mental health conditions
- (iv) Information about data completeness

All data presented in this profile were from financial year 2017/2018.

### *Population demographics*

Demographic information for these populations included age, sex, ethnicity, and deprivation index. There was also information relating to body mass index (BMI), and smoking status. Information about age, sex, and ethnicity came from primary care records. Patients were classified into nine ethnic groups: white, or not stated; Indian; Pakistani; Bangladeshi; other Asian; Black Caribbean; Black African; Chinese; other ethnic group. Deprivation was measured using the Index of Deprivation. This measure is provided for LSOAs – small areas that are focused on social homogeneity and a defined population size (1000 – 1500 residents) [10]. The ELHCP dataset allocates each LSOA in Tower Hamlets and City & Hackney to a decile of deprivation with ‘1’ being the most deprived decile. Information relating to BMI and smoking status came from primary care records.

### *Mental and physical health conditions*

Primary care data were used to identify and define nine mental health conditions and 15 physical health conditions. These conditions are listed in Table 1. Variables pertaining to each condition were formulated by NEL CSU using EMIS Read Codes (v.2 and v.3). Lists of the codes used to formulate each variable are available on request. All physical health variables were deemed long-term conditions and therefore a patient who had ever had a primary care diagnosis recorded were included in the physical health cohort. This was the same for severe

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

mental illness (SMI, i.e. schizophrenia, psychosis, bipolar disorder), personality disorder, anorexia, bulimia, dementia, and learning disabilities. As depression, low mood, and anxiety can potentially be more episodic conditions, further work was done to determine the timelines which should be used to define these cohorts. Based on local clinical and academic advice, and available evidence [11] we decided to create two patient cohorts for each conditions – patients who had ever received a diagnosis and patients who had a diagnosis recorded in the last 12 months.

Looking at the impact of the method of definition on activity in secondary care services indicated that the preferable way to define the depression, low mood, and anxiety cohorts was to include those who had ever had a diagnosis recorded in primary care. This was partly because only looking at those who had received a diagnosis in the past 12 months meant that we omitted a considerable proportion of patients who likely had common mental health disorders (CMDs, i.e. depression and/or anxiety) from our analyses. If required, it is possible to apply defined periods of time to conditions where full prolonged remission or cure is a possibility, e.g. cancer.

Prevalence rates for each condition in both Tower Hamlets and City & Hackney were calculated from the data. We assessed population prevalence in the dataset and compared with both the local and national population from other data sources. The prevalence rates of all 15 physical health conditions across six of the mental health conditions were also calculated in order to illustrate the bidirectional links between mental and physical health.

*Health service utilisation*

Health service utilisation data outlined in this profile covered use of emergency services for exemplary purposes. Data relating to specific mental health conditions were described. Health service utilisation data for elective and non-elective inpatient admissions, and outpatient appointments are also available in the ELHCP Data Repository. These data come from the SUS dataset and provide number of attendances/admissions/appointments per patient per year. There is also data pertaining to associated spend (e.g. spend per activity, spend per patient per year) which is currently being validated and is therefore not presented in this paper.

*Data completeness*

As the ELHCP Data Repository comprises routinely collected data from several different data flows there is considerable variation in completeness across variables. Completeness of certain variables appears to be better in patients with diagnoses of mental/physical health conditions

compared to those with no diagnoses. We used type 2 diabetes and COPD as exemplar physical conditions, and depression and SMI as exemplar mental conditions to demonstrate how data quality differs across patient groups. We provided completeness counts on several demographic and clinical variables including age, sex, deprivation level, ethnicity, smoking status, alcohol use, BMI, weight (kg), systolic blood pressure (SBP), diastolic blood pressure (DBP), and cholesterol. All variables were from primary care datasets. Alcohol use was measured using a composite variable of combined scores from the Alcohol Use Disorders Identification Test (AUDIT) and AUDIT-consumption (AUDIT-C) completed in primary care.

## **Cohort description and findings to date**

### *Population demographics*

The GP-registered populations for Tower Hamlets and City & Hackney in financial year 2017/2018 were 351,749 and 344,511, respectively. See Table 2 for a summary of population demographics for both Tower Hamlets and City & Hackney. Data for age, sex, and smoking status were recorded for the complete GP-registered populations in both East London boroughs. Deprivation index was recorded for most of the population (99.8% complete in both boroughs). Ethnicity data was recorded to a good level of completeness at 86.6% and 84.6% in Tower Hamlets and City & Hackney respectively. BMI data was more problematic and was recorded in the primary care records for less than a quarter of GP-registered patients (Tower Hamlets: 23.4%, City & Hackney: 24.0%).

Demographic data suggested that both Tower Hamlets and City & Hackney are ‘young’ boroughs relative to English national data[12] (median age 39 years) with mean ages of 32 (median = 31 years) and 34 years (median = 33 years) respectively. In terms of ethnicity, the majority of the GP-registered population in Tower Hamlets (38.2%) and City & Hackney (46.8%) were of white (or not stated) ethnicity. A considerable proportion of the population of Tower Hamlets were of Bangladeshi ethnicity (27.4%) and a significant amount of the City & Hackney population belonged to ‘other ethnic group’ (16.8%). Most of the GP-registered population in both East London boroughs were living in areas of high deprivation: in Tower Hamlets and City & Hackney 58.8% and 54.2% of the population respectively were in deciles 1 and 2, indicative of high deprivation. Approximately 20% of the population were recorded in primary care as being smokers. Although BMI is poorly recorded in primary care records, mean BMI scores indicated that the GP-registered population on average were overweight (Tower Hamlets: 26.3±8.1; City & Hackney: 27.1±8.6).

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

*Prevalence rates of mental and physical health conditions*

Prevalence rates for GP-coded diagnoses of mental health conditions in both Tower Hamlets and City & Hackney are provided in Figure 2. Prevalence rates were similar across both boroughs for all conditions. Comparative local and national prevalence data for each condition is provided in Table 3. We did not provide age- and sex-standardised prevalence rates from the ELHCP Data Repository to allow direct comparison with unstandardized local prevalence data from PHE. The GP-coded prevalence of common mental health disorders (CMDs) was 16.8% and 17.5% in Tower Hamlets and City & Hackney respectively. In this case, CMD comprised diagnoses of depression, anxiety, as well as low mood. Low mood is often recorded both in place of and alongside depression in primary care and is therefore important to consider in a measure of CMD. For example, in Tower Hamlets of those with a diagnosis of depression 33% also had a diagnosis of low mood. CMD prevalence was in line with data from the latest Adult Psychiatric Morbidity Survey (APMS) which reported that 17% of the English population have a CMD[13]. However, the GP-coded prevalence of depression in Tower Hamlets (8.7%) and City & Hackney (11.4%) was higher than Public Health England (PHE) prevalence data[14] from the same areas from the same financial year (7.3%, 9.6%). This could be due to the GP-coded prevalence in this profile reflecting diagnoses of depression that were ever recorded, which likely led to somewhat inflated rates.

GP-coded prevalence of SMI (1.3%) from the ELHCP Data Repository was in line with data from the PHE Psychosis Data Report from 2016[9] which reported prevalence rates of SMI for both Tower Hamlets and City & Hackney at 1.32% and 1.36% respectively. Prevalence rates for learning disability were also in line with local and national data provided by Public Health England[14]. However, there were discrepancies between the GP-coded prevalence of personality disorder, eating disorders, and dementia in the ELHCP Data Repository and local and national PHE[14] and APMS[13] data. What this suggests is that these conditions might be under-recorded in primary care in these boroughs.

Prevalence rates for GP-coded diagnoses of physical health conditions in the ELHCP Data Repository are provided in Figure 3. Prevalence rates were similar across boroughs for most conditions. However, primary care records showed that levels of type 2 diabetes were higher in Tower Hamlets and levels of hypertension were higher in City & Hackney. Comparative local and national prevalence data for each condition is provided in Table 4. Local and national prevalence data was sourced from PHE public health profiles[14]. Overall, GP-coded

prevalence of physical health conditions from the ELHCP Data Repository seems to be in line with PHE local data for the same areas. There is one notable discrepancy, however. Levels of asthma were much higher in the ELHCP Data Repository. This might be due to overdiagnosis of asthma in primary care which has been reported in both adults[15] and children[16]. Moreover, PHE prevalence data is taken directly from Quality Outcomes Framework (QOF) registers. Prevalence data from QOF registers might differ from data from other sources due to coding or definitional issues, as well as population churn. Within the ELHCP Data Repository 'liver disease' comprises many liver complications, e.g. hepatitis, fatty liver disease, alcoholic liver disease. Therefore, it was not possible to directly compare prevalence data relating to liver disease with local and national statistics as PHE provide individual prevalence rates for each liver disease subtype.

#### *Prevalence of physical health conditions among mental health conditions*

Mental and physical health are inextricably linked[17]. Having a mental health condition predisposes an individual to develop problems with their physical health. At the same time, living with a physical health condition can greatly impact mental health. Figure 4a and Figure 4b illustrate the prevalence of physical health conditions among mental health conditions in the ELHCP Data Repository, obtained from the N. Mental conditions included in this analysis were depression, low mood, anxiety, SMI, and personality disorder.

Results indicate that levels of all 15 physical health conditions were increased in those with mental health conditions. The distribution of the prevalence of physical health conditions amongst mental health conditions was similar across boroughs. The most pronounced increases in physical health conditions were seen for those flagged by their GP as having depression, SMI, and personality disorder. Some striking findings included the increase in levels of type 2 diabetes and hypertension among patients flagged by their GP as having SMI. Type 2 diabetes is known to be elevated in patients with SMI[18,19]. This increase might be brought about through use of antipsychotic medications known to be both diabetogenic and linked with weight gain[20], as well as through social deprivation, poor nutrition, chronic stress, and lifestyle changes[21]. Associations between SMI and hypertension reported in the literature have been more mixed[21]. Another striking finding that occurred in both boroughs was the high prevalence of asthma for people with a primary care diagnosis of personality disorder. This may be linked to smoking which is known to be increased in patients with personality disorder[22]. Personality disorder has been linked to hypochondriasis and perceived

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

breathlessness[23], which might also be contributing factors to higher levels of asthma diagnosis in primary care for these patients.

*A&E attendances across mental health conditions*

We used A&E as an example of a health service utilisation outcome measure to illustrate some of the activity data available in the ELHCP Data Repository. Data relating to use of A&E across both boroughs are detailed in Table 5. Based on crude attendance rates (per 1000 population), patients with mental health conditions attended A&E 54% and 76% more in Tower Hamlets and City & Hackney respectively compared to those without GP-recorded mental health conditions. Increased attendance rates were particularly pronounced in patients with dementia, personality disorder, and SMI. Similar health service utilisation data exists relating to inpatient admissions and use of outpatient services.

*Data completeness*

The ELHCP Data Repository relies on data inputted by health care provider organisations (GP practices, clinical coders in acute care settings, etc.). It is comprised of routinely collected data that can vary in completeness across variables. Examination of four example physical and mental health conditions illustrated how data completeness varies across patient groups. Table 6 details completeness on key demographic and clinical variables for patients with type 2 diabetes, COPD, depression, and SMI in Tower Hamlets. These data were all recorded in primary care databases. Data pertaining to age, sex, deprivation level, ethnicity, and smoking status were complete for all patients included in the analysis. Data relating to alcohol use and clinical factors differed between those with and without health conditions in that the data was considerably more complete for patients registered as having a condition. This improvement in completeness is likely down to the implementation of financial incentive schemes introduced as part of the UK’s QOF to improve clinical management of patients with chronic conditions[4,24]. What this illustrates is that the ELHCP Data Repository is well-suited to perform research on individual physical and mental health conditions using primary care data, where data completeness is to a high standard.

*Patient and public involvement*

The primary purpose of this project was for the health and care sector to use for system planning and effective commissioning, and therefore public stakeholders in this case comprised primary care managers, commissioners (GPs), public health experts and clinicians in primary and



secondary care. A clinical expert group was convened consisting of fifteen clinicians, commissioners and public health experts from primary care and secondary care (mental health) services. This group met on a quarterly basis during initial profiling of the cohort and provided feedback into the analysis and emerging clinical and demographic characteristics.

### *Data security*

The dataset contains sensitive patient level data; therefore, it was and remains important to ensure appropriate protections are in place. All data are de-identified, with no information such as name, home address or NHS number. Also, certain sensitive health data such as records relating to trauma and abuse, or specific sexual health data, are not visible.

To further preserve anonymity and to ensure proportionate and appropriate use of the data, access to records are limited to only being able to view aggregated records where there is a minimum of 5 individuals. Searches where there are fewer than 5 records are suppressed. Access to the complete de-identified records is further limited to specific individuals who are employed by, or under contract with, the health and care sector, have knowledge of how to use the data effectively and have undergone appropriate training.

This dataset has been compiled for the health and care sector to use for system planning. Currently plans are being made to deploy the data into an electronic platform for internal use: recognising the possible value this data holds for researchers, there may be scope to support analysis deemed to be in the public interest, with appropriate data protections in place. Currently, permissions must be sought on a case by case basis and approved by relevant IG governance groups.

## **Strengths and limitations**

### *Strengths*

One of the main strengths of the dataset is that it provides the opportunity to analyse a relatively young, urban, ethnically diverse, and considerably deprived population: groups that are hard to reach with traditional survey methods. Given patient records are linked across care settings, the data can be used to evaluate the impacts of integrated interventions. In this profile we have illustrated that the ELHCP Data Repository is representative in terms of the prevalence of most mental and physical health conditions compared to other sources of local and national

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

prevalence data supporting its suitability for real-world research and decision-making. This dataset also allows for the analysis of comorbidities and multimorbidity. Moreover, there is a large amount of data available around health service utilisation and associated spend meaning that the dataset can be used to understand what drives use and spend and support healthcare commissioning and health service delivery.

The dataset is in regular use by commissioners to better understand the health and care needs of the population, highlight where there is scope for improved patient care, and evaluate the impact of health and care interventions of patients and the system. This has provided commissioners, and others working on health and care system integration, with more sophisticated evidence and analysis to inform decisions about where to target improvement efforts to maximise benefit to both patients and the system. The findings presented in this cohort profile have highlighted where links between mental and physical health are most prevalent in the population and helped to shape commissioning priorities accordingly. For example, analysis highlighting mental and physical comorbidity supported conversations about expanding IAPT services within long term condition pathways, noting that it may be particularly impactful for people with type 2 diabetes, respiratory conditions, and cancer. Furthermore, over the next year Tower Hamlets is investing in a new 24/7 core fidelity crisis service, and the transformation of community mental health services to become embedded in primary care networks. These new models will support the assessment and treatment of patients at the right place at the right time. The ELHCP Data Repository described in this cohort profile will be an important tool in evaluating the impact these service changes have on patients and the healthcare system.

In terms of completeness, this paper highlighted that data were considerably more complete for patients with long-term mental or physical illness (i.e. type 2 diabetes, COPD, depression, SMI) compared to patients without. This supports the suitability of the dataset for the analysis of patients with long-term conditions.

*Limitations*

In terms of limitations associated with the dataset, in some cases prevalence data was not in line with local and national data, meaning that it is possible some conditions are under-recorded in primary care, i.e. personality disorder, eating disorders, and dementia. Also, as prevalence data relies on GP diagnoses, it is not possible to gauge disease duration in patients as the date in which a diagnosis is first recorded might not be the date the patient received the diagnosis.



For example, the diagnosis might have been recorded on the day the patient registered with their GP, reflecting that date rather than the date of diagnosis. In terms of data completeness, although this is good for patients with long-term conditions, there is considerable missingness among 'healthy' patients in the dataset. This has also been reported in a study of a large primary care databases in the UK which partially attributed the discrepancy in missing data to GP incentive schemes [25]. It is possible to resolve this with imputation to some degree. As is the case with all routine clinical data, the dataset requires extensive cleaning to remove duplicate entries and spurious values out of clinical range (e.g. BMI, blood pressure) prior to analysis. Another limitation involves the amount of data available. Currently, there are two financial years of data that are at an acceptable standard for research and/or decision support. New data continues to be collected and flow into this data set, so the time series available will increase over time. A further limitation surrounds the ability to measure population mobility in a dataset of this kind. One way is to look at the length of GP registration as a proxy for this, but this is problematic as it is not possible to determine whether the mobility is within the borough, or whether patients have come from outside.

### **Conclusions and future collaborations**

This cohort profile describes a unique patient-linked dataset in East London which can be used to facilitate the development of similar datasets in other parts of the UK. Moreover, this dataset provides a unique opportunity to analyse a relatively young, urban, diverse, and deprived population in East London. The dataset is growing year on year and data quality should improve over time making this an invaluable resource for researchers, clinicians, commissioners, and policy makers. The data are being further cleaned and evaluated using imputation, Bayesian, and economic methods, including engagement with machine learning work funded by the Alan Turing Institute to improve and automate the analysis of large scale electronic medical records (KNIFE project). Queries regarding data access should be directed to Katie Brennan ([katie.brennan1@nhs.net](mailto:katie.brennan1@nhs.net)) or Nathan Cheetham ([nathan.cheetham1@nhs.net](mailto:nathan.cheetham1@nhs.net)) at the NHS Tower Hamlets CCG.

### **Author statement**

AR drafted the manuscript with input and critical revisions from all authors. QK, KB, EW, GH, RF, MF, and KB provided supervision throughout. KB is the principal investigator and grant holder. Analysis of the data was performed by AA, EC, and IE. All authors reviewed the paper before submission.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Competing interest statement:** Dr Mark Freestone received research grants from East London NHS Foundation Trust and The Alan Turing Institute during the conduct of this study.

**Funding:** Funding for this project was granted by East London NHS Foundation Trust (ELFT) (184091163), the Tower Hamlets Together Board, the Health London Partnership Mental Health Transformation Board, and the City & Hackney Transformation Board. The funder had no role in the preparation of this manuscript. This report is independent research supported by the National Institute for Health Research ARC North Thames. The views expressed in this publication are those of the author(s) and not necessarily those of the National Institute for Health Research or the Department of Health and Social Care.

**Data availability statement:** The data is housed at the NHS Tower Hamlets Clinical Commissioning Group. Permission to access the data will need to be requested from Katie Brennan ([katie.brennan1@nhs.net](mailto:katie.brennan1@nhs.net)) or Nathan Cheetham ([nathan.cheetham1@nhs.net](mailto:nathan.cheetham1@nhs.net)) at the NHS Tower Hamlets Clinical Commissioning Group.

**References**

- 1 Cook JA, Collins GS. The rise of big clinical databases. *Br J Surg* 2015;**102**:e93–101. doi:10.1002/bjs.9723
- 2 Bates DW, Saria S, Ohno-Machado L, *et al.* Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Aff Proj Hope* 2014;**33**:1123–31. doi:10.1377/hlthaff.2014.0041
- 3 Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst* 2014;**2**:3. doi:10.1186/2047-2501-2-3
- 4 Correa A, Hinton W, McGovern A, *et al.* Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) sentinel network: a cohort profile. *BMJ Open* 2016;**6**:e011092. doi:10.1136/bmjopen-2016-011092
- 5 Herrett E, Gallagher AM, Bhaskaran K, *et al.* Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015;**44**:827–36. doi:10.1093/ije/dyv098
- 6 Perera G, Broadbent M, Callard F, *et al.* Cohort profile of the South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLaM BRC) Case Register: current status and recent enhancement of an Electronic Mental Health Record-derived data resource. *BMJ Open* 2016;**6**:e008721. doi:10.1136/bmjopen-2015-008721
- 7 Tower Hamlets Council. Ethnicity in Tower Hamlets: Analysis of 2011 census data. 2013.

- [https://www.towerhamlets.gov.uk/Documents/Borough\\_statistics/Ward\\_profiles/Census-2011/RB-Census2011-Ethnicity-2013-01.pdf](https://www.towerhamlets.gov.uk/Documents/Borough_statistics/Ward_profiles/Census-2011/RB-Census2011-Ethnicity-2013-01.pdf) (accessed 7 Mar 2019).
- 8 Trust for London. Poverty and inequality data for Tower Hamlets. Trust Lond. 2017./data/boroughs/tower-hamlets-poverty-and-inequality-indicators/ (accessed 3 Jul 2019).
  - 9 Public Health England. Psychosis data report. 2016;:77.
  - 10 Noble M, McLennan D, Wilkinson K, *et al*. English Indices of Deprivation. London: : Communities and Local Government 2007.
  - 11 John A, McGregor J, Fone D, *et al*. Case-finding for common mental disorders of anxiety and depression in primary care: an external validation of routinely collected data. *BMC Med Inform Decis Mak* 2016;**16**:35. doi:10.1186/s12911-016-0274-7
  - 12 Ethnicity facts and figures: Age groups. HM Government Services 2018. <https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/demographics/age-groups/latest#download-the-data> (accessed 6 Jun 2019).
  - 13 Adult Psychiatric Morbidity Survey: Survey of Mental Health and Wellbeing, England, 2014. NHS Digit. <https://digital.nhs.uk/data-and-information/publications/statistical/adult-psychiatric-morbidity-survey/adult-psychiatric-morbidity-survey-survey-of-mental-health-and-wellbeing-england-2014> (accessed 11 Jan 2019).
  - 14 Public Health England. Public Health Profiles. <https://fingertips.phe.org.uk/profile-group/mental-health> (accessed 18 Jun 2019).
  - 15 Aaron SD, Vandemheen KL, Boulet L-P, *et al*. Overdiagnosis of asthma in obese and nonobese adults. *CMAJ* 2008;**179**:1121–31. doi:10.1503/cmaj.081332
  - 16 Akker IL den, Luijn K van, Verheij T. Overdiagnosis of asthma in children in primary care: a retrospective analysis. *Br J Gen Pract* 2016;**66**:e152–7. doi:10.3399/bjgp16X683965
  - 17 Doherty AM, Gaughran F. The interface of physical and mental health. *Soc Psychiatry Psychiatr Epidemiol* 2014;**49**:673–82. doi:10.1007/s00127-014-0847-7
  - 18 Das-Munshi J, Ashworth M, Gaughran F, *et al*. Ethnicity and cardiovascular health inequalities in people with severe mental illnesses: protocol for the E-CHASM study. *Soc Psychiatry Psychiatr Epidemiol* 2016;**51**:627–38. doi:10.1007/s00127-016-1185-8
  - 19 Ward M, Druss B. The epidemiology of diabetes in psychotic disorders. *Lancet Psychiatry* 2015;**2**:431–51. doi:10.1016/S2215-0366(15)00007-3
  - 20 Gianfrancesco FD, Grogg AL, Mahmoud RA, *et al*. Differential effects of risperidone, olanzapine, clozapine, and conventional antipsychotics on type 2 diabetes: findings from a large health plan database. *J Clin Psychiatry* 2002;**63**:920–30.
  - 21 Osborn DPJ, Wright CA, Levy G, *et al*. Relative risk of diabetes, dyslipidaemia, hypertension and the metabolic syndrome in people with severe mental illnesses:

systematic review and metaanalysis. *BMC Psychiatry* 2008;**8**:84. doi:10.1186/1471-244X-8-84

22 Zvolensky MJ, Jenkins EF, Johnson KA, *et al.* Personality Disorders and Cigarette Smoking among Adults in the United States. *J Psychiatr Res* 2011;**45**:835–41. doi:10.1016/j.jpsychires.2010.11.009

23 Chetta A, Gerra G, Foresi A, *et al.* Personality Profiles and Breathlessness Perception in Outpatients with Different Gradings of Asthma. *Am J Respir Crit Care Med* 1998;**157**:116–22. doi:10.1164/ajrccm.157.1.9702093

24 Kontopantelis E, Reeves D, Valderas JM, *et al.* Recorded quality of primary care for patients with diabetes in England before and after the introduction of a financial incentive scheme: a longitudinal observational study. *BMJ Qual Saf* 2013;**22**:53–64. doi:10.1136/bmjqs-2012-001033

25 Petersen I, Welch CA, Nazareth I, *et al.* Health indicator recording in UK primary care electronic health records: key implications for handling missing data. *Clin Epidemiol* 2019;**11**:157–67. doi:10.2147/CLEP.S191437

**Table 1.** Mental and physical health conditions recorded in primary care records**Mental health conditions**

Depression; low mood\*; anxiety; severe mental illness (SMI); personality disorder; anorexia; bulimia; dementia; learning disability

**Physical health conditions**

Type 1 diabetes, type 2 diabetes, asthma, chronic obstructive pulmonary disorder (COPD), cancer (excluding basal cell carcinoma), stroke, coronary heart disease (CHD), heart failure, atrial fibrillation, hypertension, peripheral artery disease (PAD), chronic kidney disease (CKD), liver/pancreas complications (e.g. hepatitis, liver disease, chronic pancreatitis), osteoporosis, epilepsy

\*low mood is often recorded in primary care in place of depression and is therefore important to consider in any analyses relating to mental health

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Table 2.** Population demographics for the GP-registered populations of Tower Hamlets and City & Hackney

	Tower Hamlets	City & Hackney
<b>GP registered population (n,%)</b>	351749 (100%)	344511 (100%)
<b>Age (years) (M±SD)</b>	32 ± 17.97	34 ± 19.47
<b>Sex (n,%)</b>		
<i>Male</i>	178088 (50.63%)	168371 (48.87%)
<i>Female</i>	173656 (49.37%)	176134 (51.13%)
<b>Ethnicity (n,%)</b>		
<i>White or not stated</i>	134411 (38.21%)	161069 (46.75%)
<i>Indian</i>	9295 (2.64%)	9067 (2.63%)
<i>Pakistani</i>	3547 (1.01%)	2707 (0.79%)
<i>Bangladeshi</i>	96418 (27.41%)	5877 (1.71%)
<i>Other Asian</i>	8151 (2.32%)	6189 (1.80%)
<i>Black Caribbean</i>	4281 (1.22%)	16052 (4.66%)
<i>Black African</i>	14034 (3.99%)	30037 (8.72%)
<i>Chinese</i>	10435 (2.97%)	4137 (1.20%)
<i>Other ethnic group</i>	23934 (6.80%)	56140 (16.29%)
<b>Deprivation index (n,%)</b>		
<i>High deprivation*</i>	206723 (58.77%)	186784 (54.22%)
<i>Low deprivation**</i>	4955 (1.41%)	3134 (0.91%)
<b>Smokers (n,%)</b>	69627 (19.79%)	72371 (21.01%)
<b>BMI (kg/m<sup>2</sup>) (M±SD)</b>	26.34 ± 8.28	27.06 ± 8.58

\*High deprivation refers to the GP-registered population in decile 1 and 2  
\*\*Low deprivation refers to the GP-registered population in decile 9 and 10  
BMI = body mass index; M: Mean; SD: Standard deviation

Table 3. GP-coded prevalence of mental health conditions in Tower Hamlets and City & Hackney compared to local and national prevalence data (sources: APMS 2014, PHE Public Health Profiles, PHE Psychosis Data Report 2016)

Condition	GP-coded prevalence				Local prevalence		National prevalence
	<i>Tower Hamlets</i>	<i>95% CI</i>	<i>City &amp; Hackney</i>	<i>95% CI</i>	<i>Tower Hamlets</i>	<i>City &amp; Hackney</i>	<i>England</i>
<b>CMDs</b>	16.8%	16.69,16.94	17.5%	17.37,17.62	15.4%[14]	17%[14]	17%[13]
<b>SMI</b>	1.3%	1.24,1.31	1.3%	1.31,1.38	1.32%[9]	1.36%[9]	0.5-1.5%[9]
<b>Personality disorder</b>	0.5%	0.49,0.53	0.6%	0.55,0.60	-	-	13.7%[13]*
<b>Eating disorders</b>	0.3%	0.29,0.33	0.4%	0.37,0.41	6.9%[14]	7.4%[14]	6.9%[14]
<b>Dementia</b>	0.4%	0.34,0.38	0.4%	0.38,0.42	5.25%[14]	4.37%[14]	4.33%[14]
<b>Learning disability</b>	0.3%	0.32,0.36	0.4%	0.36,0.41	0.3%[14]	0.4%[14]	0.5%[14]

\*Percentage of sample who scored  $\geq 4$  on the Standardised Assessment of Personality – Abbreviated Scale (SAPAS)

APMS = Adult Psychiatric Morbidity Survey; CMDs = common mental health disorders; GP = general practitioner; PHE = Public Health England; SMI = severe mental illness

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46

**Table 4.** GP-coded prevalence of physical health conditions in Tower Hamlets and City & Hackney compared to local and national prevalence data (source: PHE Public Health Profiles)

Condition	GP-coded prevalence		Local prevalence		National prevalence
	<i>Tower Hamlets</i>	<i>City &amp; Hackney</i>	<i>Tower Hamlets</i>	<i>City &amp; Hackney</i>	<i>England</i>
Type 1 diabetes*	0.3%	0.4%			
Type 2 diabetes*	5.1%	4.4%	6.7%	5.7%	6.8%
Asthma	8.7%	9.2%	4.3%	4.4%	5.9%
COPD	1.3%	1.2%	1.2%	1.1%	1.9%
Cancer	1.4%	1.8%	1.1%	1.4%	2.7%
Stroke	0.8%	0.9%	0.9%	0.7%	1.8%
CHD	1.6%	1.7%	1.5%	1.5%	3.1%
Heart failure	0.7%	0.9%	0.4%	0.7%	0.8%
Atrial fibrillation	0.6%	0.8%	0.6%	0.7%	1.9%
Hypertension	7.4%	9.3%	7.5%	9.3%	13.9%
PAD**	0.3%	0.4%	0.96%	1.01%	1.16%
CKD	2.0%	1.8%	2.5%	2.1%	4.1%
Liver disease†	1.9%	1.4%	-	-	-
Osteoporosis	0.06%	0.09%	-	-	0.6%
Epilepsy	0.8%	0.8%	0.4%	0.5%	0.8%

\*PHE does not provide prevalence rates for diabetes subtypes. Therefore local and national prevalence rates is for diabetes overall.



\*\*Local and national prevalence data for PAD from PHE were based on modelled estimated prevalence

†It was not possible to provide local and national prevalence rates for liver disease as data is provided separately by PHE for disease subtype (e.g. hepatitis C, fatty liver disease, alcoholic liver disease)

CHD = coronary heart disease; CKD = chronic kidney disease; COPD = chronic obstructive pulmonary disorder; GP = general practitioner; PAD = peripheral artery disease; PHE = Public Health England

For peer review only

**Table 5.** The impact of mental health conditions on A&E attendances, financial year 2017/2018

Accident & Emergency								
Activity (attendances)								
	Total attendances		% Share		Crude rate (per 1000 pop.)			
	TH	C&H	TH	C&H	TH	95% CI	C&H	95% CI
Population	159201	161641	100%	100%	452	412, 495	469	428, 513
All non-mental health	118272	116816	74.3%	72.3%	412	373, 453	413	375, 454
All mental health	40929	44825	25.7%	27.7%	635	587, 686	728	677, 782
CMDs	31625	34215	19.9%	21.2%	535	491, 582	568	523, 616
SMI	3800	4383	2.4%	2.7%	847	791, 906	945	886, 1007
Personality disorder	2529	2914	1.6%	1.8%	1413	1341, 1488	1467	1393, 1543
Eating disorder	584	549	0.4%	0.3%	554	509, 602	424	385, 466
Dementia	1595	1471	1.0%	0.9%	1248	1180, 1319	1075	1012, 1141
Learning disability	796	1293	0.5%	0.8%	670	621, 722	974	914, 1037

A&E = Accident & Emergency; CMDs = common mental health disorders; C&H = City & Hackney; SMI = severe mental illness; TH = Tower Hamlets

**Table 6.** ELHLP data completeness (%) across physical and mental health conditions in Tower Hamlets, financial year 2017/2018

	<b>T2DM</b>	<b>No T2DM</b>	<b>COPD</b>	<b>No COPD</b>	<b>Depression</b>	<b>No Depression</b>	<b>SMI</b>	<b>No SMI</b>
<i>Variables</i>	<i>% complete</i>	<i>% complete</i>	<i>% complete</i>	<i>% complete</i>	<i>% complete</i>	<i>% complete</i>	<i>% complete</i>	<i>% complete</i>
<b>Age</b>	100%	100%	100%	100%	100%	100%	100%	100%
<b>Sex</b>	100%	100%	100%	100%	100%	100%	100%	100%
<b>Deprivation</b>	100%	100%	100%	100%	100%	100%	100%	100%
<b>Ethnicity</b>	100%	100%	100%	100%	100%	100%	100%	100%
<b>Smoking status</b>	100%	100%	100%	100%	100%	100%	100%	100%
<b>Alcohol use</b>	94.6%	51.1%	91.8%	52.8%	73.1%	51.4%	89.2%	52.8%
<b>BMI</b>	99.3%	72.6%	99.3%	73.7%	95.4%	71.9%	97.5%	73.7%
<b>Weight</b>	99.4%	84.6%	99.5%	85.2%	96.7%	84.3%	98.2%	85.2%
<b>SBP</b>	99.6%	68.8%	99.6%	69.9%	96.4%	67.9%	98.5%	69.9%
<b>DBP</b>	99.6%	68.7%	99.6%	69.9%	96.3%	67.9%	98.5%	69.9%
<b>Cholesterol</b>	99.4%	28.8%	96.9%	31.6%	64.9%	29.3%	84.6%	37.8%

BMI = body mass index; COPD: chronic obstructive pulmonary disorder; DBP = diastolic blood pressure; ELHLP = East London Healthcare Links Project; SBP = systolic blood pressure; SMI = severe mental illness; T2DM = Type 2 diabetes mellitus

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Figure 1.** Diagram of data flows from healthcare providers to Tower Hamlets & City Hackney CCGs

CCG = Clinical Commissioning Group, CDS = Commissioning Data Sets, CSU = Commissioning Support Unit, DCB = Data Coordination Board, DSCRO = Data Services for Commissioners Regional Office, NEL = North East London, NHS = National Health Service

**Figure 2.** GP-coded prevalence rates of mental health conditions in Tower Hamlets and City & Hackney, financial year 2017/2018.

\*CMD incorporates depression, anxiety and low mood

**Figure 3.** GP-coded prevalence rates of physical health conditions in Tower Hamlets and City & Hackney, financial year 2017/2018

**Figure 4a.** Prevalence of physical health conditions among mental health conditions in Tower Hamlets

**Figure 4b.** Prevalence of physical health conditions among mental health conditions in City & Hackney

